

Searching for the shadows of giants: characterising protoclusters with line of sight Lyman- α absorption

Joel S. A. Miller^{*}, James S. Bolton and Nina Hatch

School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK

6 September 2019

ABSTRACT

We use state of the art hydrodynamical simulations from the Sherwood, EAGLE and Illustris projects to examine the signature of $M_{z=0} \simeq 10^{14} M_{\odot}$ protoclusters observed in Ly- α absorption at $z \simeq 2.4$. We find there is a weak correlation between the mass overdensity, δ_m , and the Ly- α effective optical depth relative to the mean, $\delta_{\tau_{\text{eff}}}$, averaged over $15 h^{-1}$ cMpc scales, although scatter in the $\delta_m - \delta_{\tau_{\text{eff}}}$ plane means it is not possible to uniquely identify large scale overdensities with strong Ly- α absorption. Although all protoclusters are associated with large scale mass overdensities, most sight lines through protoclusters in a $\sim 10^6$ cMpc³ volume probe the low column density Ly- α forest. A small subset of sight lines that pass through protoclusters exhibit coherent, strong Ly- α absorption on $15 h^{-1}$ cMpc scales, although these correspond to a wide range in mass overdensity. Assuming perfect removal of contamination by Ly- α absorbers with damping wings, more than half of the remaining sight lines with $\delta_{\tau_{\text{eff}}} > 3.5$ trace protoclusters. It is furthermore possible to identify a model dependent $\delta_{\tau_{\text{eff}}}$ threshold that selects only protoclusters. However, such regions are rare: excluding absorption caused by damped systems, less than 0.1 per cent of sight lines that pass through a protocluster have $\delta_{\tau_{\text{eff}}} > 3.5$, meaning that any protocluster sample selected in this manner will also be highly incomplete. On the other hand, coherent regions of Ly- α absorption also provide a promising route for identifying and studying filamentary environments at high redshift.

Key words: galaxies: clusters: general – intergalactic medium – quasars: absorption lines

1 INTRODUCTION

Galaxy clusters at redshift $z = 0$ are the most massive virialised objects in the Universe, residing in dark matter haloes with $M_{z=0} \geq 10^{14} M_{\odot}$. At higher redshifts, the progenitors of these objects – which are referred to as protoclusters – are not yet virialised and are spread out over a scale of $\sim 20 h^{-1}$ cMpc at $z \sim 2$ (Chiang et al. 2013; Muldrew et al. 2015). Protoclusters are some of the densest structures on this scale at $z \sim 2 - 3$ when the cosmic star formation rate density is greatest, quasar activity peaked, and massive galaxies assembled the majority of their mass (Madau & Dickinson 2014). Protoclusters can therefore be used to study the effects of high density environments on galaxy formation and evolution at this important epoch, and can also be used constrain structure formation and cosmological models via their growth rate (Kravtsov & Borgani 2012).

The detection of protoclusters presents a challenge, however. Unlike mature clusters, they lack a hot X-ray emitting intra-cluster medium (Overzier 2016). Therefore, they

are most commonly found through the presence of large concentrations of galaxies. This a challenging observational task; protoclusters are large and diffuse objects, so they present only a small density enhancement over the field. Furthermore, protocluster galaxies are still forming, so they do not lie on a well defined red sequence (Gladders & Yee 2000).

Current observational techniques for finding protoclusters fall into two main categories. The first of these is to use typical galaxies as tracers. This has been performed using both photometric (Daddi et al. 2009; Chiang et al. 2014) and spectroscopic (Steidel et al. 2005; Cucciati et al. 2014; Lemaux et al. 2014; Chiang et al. 2015; Lemaux et al. 2017) surveys. The most massive protoclusters are rare (with number density $< 10^{-6}$ cMpc⁻³) and spatially extended. Consequently, in order to find protoclusters using this method, large cosmic volumes need to be surveyed. Until recently, galaxy redshift surveys that are deep enough to probe galaxies at $z \gtrsim 2$ only cover areas of $\lesssim 1$ deg², such as the COSMOS survey (Scoville et al. 2007), resulting in only a few tens of protoclusters discovered through this method over the last decade. This is about to change with ongoing and forthcoming surveys such as the Hyper Suprime-Cam Sub-

^{*} E-mail: joel.miller@nottingham.ac.uk

aru Strategic Program (HSC-SSP)¹ and the Large Synoptic Survey Telescope (LSST)²; using only a small subset of the HSC-SSP data, [Toshikawa et al. \(2018\)](#) have already discovered over 179 protocluster candidates.

The second of these methods is to focus searches around biased tracer objects, such as high redshift radio galaxies ([Le Fèvre et al. 1996](#); [Venemans et al. 2007](#); [Hatch et al. 2011](#); [Cooke et al. 2014](#)), quasars – or quasi-stellar objects (QSOs) – ([Wold et al. 2003](#); [Zheng et al. 2006](#); [Capak et al. 2011](#); [Wylezalek et al. 2013](#); [Morselli et al. 2014](#); [Adams et al. 2015](#)) and Ly- α blobs ([Bădescu et al. 2017](#); [Cai et al. 2017a](#)), which are thought to correlate well with high density regions at high redshift, and/or be the progenitors of local brightest cluster galaxies. These are also very bright objects, and can therefore be found using significantly shallower surveys. This approach greatly reduces the observational cost and it is how the majority of presently spectroscopically confirmed protoclusters at $z \gtrsim 1.3$ have been discovered. However, this comes at the expense of introducing a strong selection bias to the resulting sample.

Protoclusters are not only traced by overdensities of galaxies, however, but also by intergalactic hydrogen gas ([Adelberger et al. 2003](#); [Mukae et al. 2017](#)). The residual neutral hydrogen (H I) in this otherwise highly ionised medium can be detected in absorption in background quasar spectra through the resulting Ly- α absorption (i.e. the Ly- α forest). The idea of locating protoclusters as regions over-dense in H I was first proposed by [Francis & Hewett \(1993\)](#) with the first observational detection by [Francis et al. \(1996\)](#). Recently there has been renewed interest in this approach following the advent of large QSO surveys, such as the Baryon Oscillation Spectroscopic Survey (BOSS) ([Dawson et al. 2013](#)), the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) and the forthcoming Dark Energy Spectroscopic Instrument (DESI) ([Vargas-Magana et al. 2019](#)) and WEAVE-QSO surveys ([Pieri et al. 2016](#)).

The recent work has been approached from two different directions. The first is Ly- α forest tomography ([Lee et al. 2014](#); [Stark et al. 2015](#); [Lee et al. 2018](#)), where the underlying 3-D mass distribution is reconstructed using multiple Ly- α forest spectra in the same patch of sky. Applying insight obtained from a $256^3 h^{-3} \text{cMpc}^3$ collisionless dark matter simulation, [Lee et al. \(2016\)](#) have used this method to successfully detect a galaxy overdensity at $z = 2.44$ in the COSMOS field, as well as cosmic voids at $z \sim 2.3$ ([Krolewski et al. 2018](#)). The second method – which we focus on in this work – is to search for the most massive overdensities at $z = 2-4$ from a large survey volume using strong, coherent Ly- α absorption along the line of sight ([Cai et al. 2016](#)). This approach is more effective at redshifts where the density of bright background sources is too low for tomography. Guided by mock spectra extracted from a $1 h^{-3} \text{cGpc}^3$ collisionless dark matter simulation combined with an approximate scheme for modelling Ly- α absorption ([Peirani et al. 2014](#)), [Cai et al. \(2016\)](#) identified protoclusters as being closely associated with what they call Coherently Strong intergalactic Ly- α absorption systems (CoSLAs). Groups of

CoSLAs have been used to discover a massive overdensity at $z = 2.32$ ([Cai et al. 2017b](#)).

These approaches have had compelling success in locating potential protoclusters at $z \simeq 2-3$. However, fully hydrodynamical simulations that directly connect the gas distribution at $z > 2$ to present-day clusters are required to establish whether coherent Ly- α absorption accurately tracks the progenitors of the largest collapsed clusters at $z = 0$, or if a more complex relationship between mass and Ly- α absorption may disrupt this picture. Furthermore, for protocluster searches that employ individual sight lines to detect CoSLAs, high column density Ly- α absorbers that possess large damping wings can be a significant contaminant (cf. [Cai et al. 2016](#)).

In this work we shall address these points by investigating the properties of Ly- α absorption systems within protoclusters at $z > 2.4$ using fully hydrodynamical simulations performed by the Sherwood ([Bolton et al. 2017](#)), EAGLE ([Schaye et al. 2015](#)) and Illustris ([Vogelsberger et al. 2014](#)) projects. With a typical volume of order 10^6cMpc^3 , these simulations only contain of order ~ 10 clusters with masses $M_{z=0} \geq 10^{14} M_{\odot}$. On the other hand, unlike larger collisionless dark matter simulations, they include a full treatment of the gas physics using a variety of different (sub-grid) feedback models. More importantly, however, these simulations also have sufficient resolution to correctly incorporate Ly- α absorption features over a wide range of H I column densities, including absorption systems with damping wings. Our approach therefore provides a complementary perspective on the relationship between Ly- α absorption on small scales and the distribution of mass in protoclusters at $z \simeq 2.4$. Furthermore, as we explicitly track the formation of structure to $z = 0$ in the simulations, we are able to assess the completeness and contamination of a sample of protocluster candidates selected using coherent line of sight Ly- α absorption.

This paper is structured as follows. We first introduce the hydrodynamical simulations we use in Section 2, and compare the H I column density distribution in each model to observational constraints in Section 3. In Section 4 we describe the basic properties of protoclusters in the simulations. We characterise mass overdensities on large scales by their Ly- α absorption in Section 5, and examine their connection to protoclusters in Section 6. Finally, the effectiveness of using line of sight Ly- α absorption to detect candidate protoclusters at $z \simeq 2.4$ is discussed in Section 7, before we summarise and conclude in Section 8. Throughout the paper, we refer to comoving distance units using the prefix “c”.

2 HYDRODYNAMICAL SIMULATIONS

The hydrodynamical simulations used in this work are summarised in Table 1. Outputs at two different redshifts were used for each model: the output closest³ to $z = 2.4$, along with the corresponding output at $z = 0$. Below, we briefly describe the relevant properties of each simulation in turn,

¹ <https://hsc.mtk.nao.ac.jp/ssp/>

² <https://www.lsst.org>

³ This corresponds to $z = 2.4$ for Sherwood, $z = 2.478$ for EAGLE and $z = 2.44$ for Illustris.

although further, extensive descriptions of these simulations may be found elsewhere.

2.1 Sherwood

The Sherwood project (Bolton et al. 2017) consists of a set of large, high resolution simulations of the Ly- α forest performed using a modified version of the smoothed particle hydrodynamics (SPH) code P-GADGET-3, last described in Springel (2005). In this work we predominantly use the 80-1024-ps13 simulation from Bolton et al. (2017), which we shall refer to as “Sherwood”. This simulation was performed in a $80^3 h^{-3} \text{ cMpc}^3$ volume using the star formation and galactic outflow model developed by Puchwein & Springel (2013). This assumes a Chabrier (2003) initial mass function (IMF) and a wind velocity, v_w , that is proportional to the galaxy escape velocity, such that the wind mass-loading scales as v_w^{-2} .

The ultraviolet (UV) background follows the spatially uniform Haardt & Madau (2012) model, which quickly reionises the IGM at $z = 15$. This model assumes the hydrogen is optically thin and in photo-ionisation equilibrium, and has an H I photo-ionisation rate of $\Gamma_{\text{HI}} = 9.6 \times 10^{-13} \text{ s}^{-1}$ at $z = 2.4$. Additionally, a small boost to the He II photo-heating rate, $\epsilon_{\text{HeII}} = 1.7 \epsilon_{\text{HeII}}^{\text{HM12}}$, has been applied at $2.2 < z < 3.4$ to better match observational constraints on the IGM temperature during and after He II reionisation (Becker et al. 2011). The Sherwood models adopt a Planck 2013 consistent cosmology (Planck Collaboration et al. 2014) with $\Omega_m = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_b = 0.0481$, $\sigma_8 = 0.826$, $n_s = 0.963$ and $h = 0.678$.

In addition to the fiducial simulation described above, we also use the 80-1024 simulation described in Bolton et al. (2017). This does not follow a physically motivated star formation model, but is identical to our fiducial run in all other respects. Instead, gas particles with density $\Delta = \rho/\langle\rho\rangle > 1000$ and temperature $T < 10^5 \text{ K}$ are converted directly into collisionless star particles. This “quick Ly- α ” approach increases computational speed while having a minimal effect on the low column density absorbers in the Ly- α forest (Viel et al. 2004). As this removes all the cold, dense gas in the simulation, the incidence of H I absorbers with column densities $N_{\text{HI}} \gtrsim 10^{17} \text{ cm}^{-2}$ will be underpredicted. We include this approach here, however, as it will more closely approximate results from earlier work using post-processed dark matter simulations that have insufficient resolution for modelling dense gas. We refer to this model as Sherwood “QLy- α ”.

2.2 EAGLE

The EAGLE (Evolution and Assembly of GaLaxies and their Environments) simulation (Schaye et al. 2015; Crain et al. 2015; McAlpine et al. 2016) was performed with a customised version of P-GADGET-3, where the standard SPH approach has been modified following the ANARCHY scheme described by Schaye et al. (2015). The simulation we use in this work is the Ref-L0100N1504 model. This has a smaller box size than Sherwood, corresponding to $67.77^3 h^{-3} \text{ cMpc}^3$, but with a factor of $\simeq 5$ better mass resolution (see Table 1). Star formation is modelled using the approach of Schaye &

Dalla Vecchia (2008) assuming a Chabrier IMF, while stellar feedback follows the stochastic, thermal scheme described in Dalla Vecchia & Schaye (2012). In addition, EAGLE follows gas accretion onto black holes; feedback from active galactic nuclei (AGN) is included using the methodology of Booth & Schaye (2009).

The spatially uniform, optically thin UV background used in EAGLE follows Haardt & Madau (2001); this quickly reionises the hydrogen at $z = 9$. An additional 2 eV per proton of energy from He II reionisation is also added at $z \simeq 3.5$, resulting in gas temperatures at mean density that are $\sim 4000 \text{ K}$ larger compared to Sherwood by $z = 2.4$. The H I photo-ionisation rate at $z = 2.478$ in the Haardt & Madau (2001) model is $\Gamma_{\text{HI}} = 1.4 \times 10^{-12} \text{ s}^{-1}$, a factor of 1.5 larger than the more recent Haardt & Madau (2012) UV background used in Sherwood. EAGLE assumes a Λ CDM cosmology consistent with Planck Collaboration et al. (2014), where $\Omega_m = 0.307$, $\Omega_\Lambda = 0.693$, $\Omega_b = 0.04825$, $\sigma_8 = 0.8288$, $n_s = 0.9611$ and $h = 0.677$.

2.3 Illustris

Finally, we also use the Illustris-1 simulation (hereafter referred to as “Illustris”) in our analysis (Vogelsberger et al. 2014; Nelson et al. 2015). Unlike Sherwood and EAGLE, Illustris is performed with the moving-mesh hydrodynamics code AREPO (Springel 2010). Illustris has a slightly smaller volume than Sherwood ($75^3 h^{-3} \text{ cMpc}^3$), but it has the highest mass resolution of the three simulations we consider. The star formation, stellar feedback and AGN feedback models are described in detail by Vogelsberger et al. (2013). Star formation in Illustris also uses a Chabrier IMF and is based upon the Springel & Hernquist (2003) model. The stellar feedback uses a variable winds approach, where the wind velocity, v_w , is scaled to the local dark matter velocity dispersion.

The spatially uniform UV background follows Faucher-Giguère et al. (2009), which quickly reionises the hydrogen in the simulation at $z = 10.5$. This UV background model has a H I photo-ionisation rate $\Gamma_{\text{HI}} = 6.1 \times 10^{-13} \text{ s}^{-1}$ at $z = 2.44$ – a factor of 0.6 smaller than Haardt & Madau (2012) – and produces gas temperatures at mean density around 1500 K lower at $z = 2.4$ compared to Sherwood.

Furthermore, unlike Sherwood and EAGLE, Illustris uses an on-the-fly prescription for the self-shielding of hydrogen from Lyman continuum photons, following the approach of Rahmati et al. (2013). As we will discuss below, incorporating self-shielding is necessary for correctly capturing the incidence of absorption systems with column densities $N_{\text{HI}} \geq 10^{17.2} \text{ cm}^{-2}$ (i.e. absorption systems that are optically thick to Lyman continuum photons). The Illustris simulations assume a WMAP-9 consistent cosmology (Hinshaw et al. 2013), with $\Omega_m = 0.2726$, $\Omega_\Lambda = 0.7274$, $\Omega_b = 0.0456$, $\sigma_8 = 0.809$, $n_s = 0.963$ and $h = 0.704$.

2.4 Generation of mock Ly- α absorption sight-lines

Mock Ly- α absorption spectra were extracted along sight lines drawn from the Sherwood simulation using the SPH interpolation scheme described by Theuns et al. (1998), com-

Table 1. Hydrodynamical simulations used in this work. The columns list, from left to right: the simulation name, the box size in h^{-1} cMpc, the total number of resolution elements, the dark matter and gas particle masses (or equivalently for Illustris, the typical hydrodynamical cell mass), and the number of objects that have a total mass at $z = 0$ (as defined by a friends-of-friends halo finder) in the range $M_{z=0} \geq 10^{14} M_{\odot}$ (clusters), $10^{13.75} \leq M_{z=0}/M_{\odot} < 10^{14}$ (large groups) and $10^{13.5} \leq M_{z=0}/M_{\odot} < 10^{13.75}$ (small groups).

Name	Box size [h^{-1} cMpc]	N_{tot}	M_{dm} [M_{\odot}]	M_{gas} [M_{\odot}]	$M_{z=0} \geq 10^{14} M_{\odot}$ Clusters	$10^{13.75} \leq \frac{M_{z=0}}{M_{\odot}} < 10^{14}$ Large groups	$10^{13.5} \leq \frac{M_{z=0}}{M_{\odot}} < 10^{13.75}$ Small groups
Sherwood	80	2×1024^3	5.07×10^7	9.41×10^6	29	23	46
EAGLE	67.77	2×1504^3	9.70×10^6	1.81×10^6	10	15	33
Illustris	75	3×1820^3	6.26×10^6	1.26×10^6	14	18	29

bin with the Voigt profile approximation from [Tepper-García \(2006\)](#). Each sight line consists of 1024 pixels, and is drawn in a direction parallel to the simulation boundaries, starting from a position selected at random on the projection axis. A total of 30,000 sight lines (10,000 along each projection axis) were extracted from Sherwood, corresponding to an average transverse separation of $0.8h^{-1}$ cMpc. The transmitted flux in each pixel is given by $F = e^{-\tau}$, where τ is the Ly- α optical depth.

Sight lines were extracted with the same average transverse separation from EAGLE and Illustris, although there are some small differences in the methodology due to the different hydrodynamics schemes employed by these models. For EAGLE, the M_4 cubic spline kernel ([Monaghan & Lattanzio 1985](#)) used in the standard version of P-GADGET-3 was replaced with the C_2 [Wendland \(1995\)](#) kernel when performing the SPH interpolation. In Illustris there are no smoothing lengths, h_i , associated with the hydrodynamic cells. Instead, we assign these based on the volume, V_i , of each Voronoi cell, where

$$h_i = \left(\frac{3N_{\text{sph}}V_i}{4\pi} \right)^{1/3}, \quad (1)$$

and we adopt $N_{\text{sph}} = 64$ for the number of smoothing neighbours. We furthermore set all star-forming hydrogen gas with $n_{\text{H}} > 0.13 \text{ cm}^{-3}$ to be fully neutral in Illustris, correcting for the unphysical neutral hydrogen fractions produced by the sub-grid star formation model. In addition, since Illustris already incorporates self-shielded hydrogen on-the-fly, the [Rahmati et al. \(2013\)](#) prescription for self-shielding was applied in post-processing to both Sherwood and EAGLE.

Lastly, in order to correct for the approximately factor of two uncertainty in the UV background H I photoionisation rate Γ_{HI} ([Bolton et al. 2005](#)), we rescale the optical depths in each pixel of the mock spectra by a constant, such that the Ly- α forest effective optical depth obtained from all the mock sight lines, $\tau_{\text{eff}} = -\ln \langle F \rangle$, where $\langle F \rangle$ is the mean transmitted flux, matches observational constraints ([Theuns et al. 1998](#); [Lukić et al. 2014](#)). We use the τ_{eff} measurements from [Becker et al. \(2013\)](#) for this purpose. At $z = 2.4$, these data correspond to $\tau_{\text{eff}} = 0.20$.

3 THE H I COLUMN DENSITY DISTRIBUTION FUNCTION

Before proceeding to analyse the properties of protoclusters in Ly- α absorption, we must first verify if the simulations reproduce the observed distribution of H I column densi-

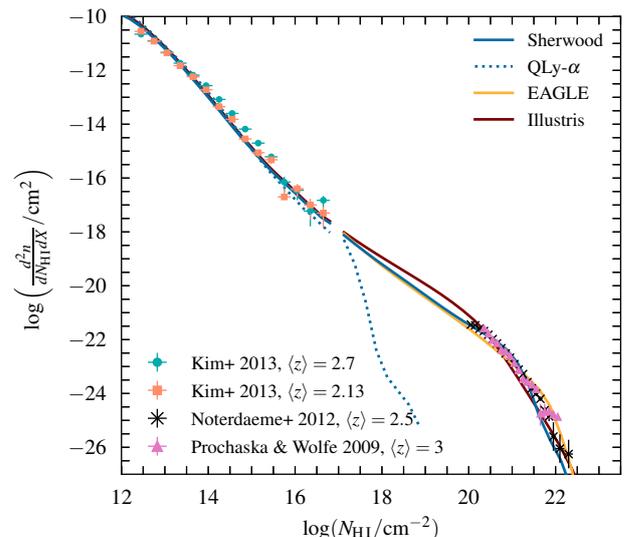


Figure 1. The H I column density distribution function (CDDF) at $z \simeq 2.4$ obtained from Sherwood (solid blue curve), EAGLE (orange curve) and Illustris (brown curve). For comparison, the dotted blue line represents the CDDF from the QLy- α simulation. Observational data from [Kim et al. \(2013\)](#) at $\langle z \rangle = 2.7$ and $\langle z \rangle = 2.13$ have been added at $N_{\text{HI}} < 10^{17} \text{ cm}^{-2}$, while data points from [Noterdaeme et al. \(2012\)](#) at $\langle z \rangle = 2.5$ and [Prochaska & Wolfe \(2009\) \$\langle z \rangle = 3\$ are displayed at \$N_{\text{HI}} > 10^{20} \text{ cm}^{-2}\$.](#)

ties⁴ at $z \simeq 2.4$. The column density distribution function (CDDF) obtained from the three simulations are displayed [Figure 1](#), along with observational measurements from [Kim et al. \(2013\)](#), [Noterdaeme et al. \(2012\)](#) and [Prochaska & Wolfe \(2009\)](#). The simulated CDDFs at $N_{\text{HI}} < 10^{17} \text{ cm}^{-2}$ are calculated by integrating the H I number density in each pixel in the mock sight lines over 50 km s^{-1} windows ([Gurvich et al. 2017](#)). However, as absorption systems with $N_{\text{HI}} > 10^{17} \text{ cm}^{-2}$ are comparatively rare, we instead compute the CDDF by projecting the H I density for the entire simulation box onto a 2-D grid consisting of 20000^2 pixels ([Altay et al. 2011](#); [Rahmati et al. 2013](#); [Bird et al. 2014](#); [Villaescusa-Navarro et al. 2018](#)). The discontinuity seen at

⁴ We will refer to Ly- α absorbers in four groups based on their column densities: Ly- α forest ($N_{\text{HI}} < 10^{17.2} \text{ cm}^{-2}$), Lyman-limit systems (LLSs, $10^{17.2} \leq N_{\text{HI}}/\text{cm}^{-2} < 10^{19}$), super Lyman-limit systems (SLLSs, $10^{19} \leq N_{\text{HI}}/\text{cm}^{-2} < 10^{20.3}$) and damped Ly- α absorbers (DLAs, $N_{\text{HI}} \geq 10^{20.3} \text{ cm}^{-2}$). We will also refer to SLLSs and DLAs collectively as “damped systems”.

$N_{\text{HI}} = 10^{17} \text{ cm}^{-2}$ in Figure 1 represents the point we switch from the integration method to the projection method.

The fiducial Sherwood simulation (solid blue curve), as well as the EAGLE (orange curve) and Illustris (brown curve) are in good agreement with observational data over a wide range of column densities, although the level of agreement at $N_{\text{HI}} > 10^{21.3} \text{ cm}^{-2}$ is possibly fortuitous given that a treatment of molecular hydrogen is not included in our analysis (Altay et al. 2013; Crain et al. 2017). The Illustris simulation also predicts a greater incidence of systems with $10^{17} \leq N_{\text{HI}}/\text{cm}^{-2} \leq 10^{20}$, relative to EAGLE and Sherwood. The reason for this difference is unclear, although we speculate this may be in part because the self-shielding correction is included on-the-fly in Illustris. The photo-heating (and hence pressure smoothing) experienced by the high column density, self-shielded gas will therefore differ from the post-processed EAGLE and Sherwood runs.

For comparison, we also show the CDDF from the QLy- α model in Figure 1 (blue dotted curve). The QLy- α model is only a good match to the observational data for the Ly- α forest at $N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$. This is because most of the overdense gas that forms the strongest absorption systems has been converted into collisionless star particles (see Section 2.1). Overall, this comparison demonstrates that Sherwood, EAGLE and Illustris will adequately capture the incidence of Ly- α absorbers at $z \simeq 2.4$ over a wide range of H I column densities.

4 SIMULATED PROTOCLUSTERS

4.1 Protocluster definitions

We now turn to describe the protoclusters in the simulations at $z \sim 2.4$. We identify the protoclusters by tracking all particles in a friends-of-friends (FoF) group⁵ at $z = 0$. We will refer to the total mass of the $z = 0$ FoF group as the mass of the protocluster, $M_{z=0}$. Earlier studies (e.g. Muldrew et al. 2015) have instead identified simulated protoclusters by tracing the merger tree of $z = 0$ haloes back to the redshift of interest. However, as we are primarily interested in following large scale Ly- α absorption, in this work we also choose to follow the mass that is not bound in haloes at $z \simeq 2.4$.

We consider three mass bins in our analysis: $M_{z=0} \geq 10^{14} M_{\odot}$ (clusters), $10^{13.75} M_{\odot} \leq M_{z=0} < 10^{14} M_{\odot}$ (large groups) and $10^{13.5} M_{\odot} \leq M_{z=0} < 10^{13.75} M_{\odot}$ (small groups). We refer to the cluster progenitors as protoclusters, and the group progenitors as protogroups; we include the latter to provide a comparison to lower mass systems. The number of FoF groups in each bin is summarised in Table 1. We also define the size of each of the protoclusters/groups at $z \simeq 2.4$ using R_{95} , which corresponds to the radius of a sphere around the protocluster centre of mass that contains 95 per cent of $M_{z=0}$.

⁵ Note this means that our quoted protocluster masses will be systematically larger than virial mass estimators such as M_{200} – the total mass enclosed within a sphere whose mean density is 200 times the critical density. For example, White (2000) demonstrate that FoF halo masses will be approximately 10 per cent greater than M_{200} .

Additionally, following an analysis of the Millennium simulation (Springel et al. 2005) by Lovell et al. (2018), for each protocluster we calculate two parameters derived from the principal semi-axes of a triaxial mass distribution, where $a \geq b \geq c$. The first is the axis ratio $s = c/a$ which provides a measure of sphericity, with $s = 1$ corresponding to a spherical distribution and $s \sim 0$ corresponding to a highly aspherical distribution. The second is the triaxiality parameter (Franx et al. 1991) given by

$$T = \frac{a^2 - b^2}{a^2 - c^2} \quad (2)$$

This quantifies whether the mass distribution resembles a prolate ($T \sim 1$) or oblate ($T \sim 0$) spheroid.

4.2 The size and shape of simulated protoclusters

The R_{95} distributions in the three mass bins are shown in the top row of Figure 2 for each simulation. The protoclusters tend to be larger than the protogroups, and typically have $R_{95} = 5\text{--}10 \text{ h}^{-1} \text{ cMpc}$. These sizes are broadly consistent with the 90 per cent stellar mass radii recovered from the Millennium simulation at $z = 2$ by Muldrew et al. (2015), as well as the radial extent determined by Hatch et al. (2011) for overdensities around radio galaxies at $z \sim 2.4$. The properties displayed in Figure 2 are also largely consistent between the different simulations, suggesting that (as expected on large scales) variations in the hydrodynamics schemes and sub-grid physics have little impact on the overall size and shape of the protoclusters. Furthermore, in all three cases, the smaller, lower mass protogroups do not differ significantly in shape from the protoclusters.

As noted by Lovell et al. (2018), however, a simple triaxial model does not fully capture the distribution of groups and filaments within protoclusters. In Figure 3, we therefore display two-dimensional projections of protoclusters selected from each of the three simulations: these have masses $M_{z=0} = 10^{14.3} M_{\odot}$ (Sherwood), $M_{z=0} = 10^{14.4} M_{\odot}$ (EAGLE) and $M_{z=0} = 10^{14.3} M_{\odot}$ (Illustris). Several different protocluster morphologies are apparent, ranging from a structure dominated by a massive central halo (Sherwood) to a more diffuse structure with multiple, lower mass haloes (e.g. EAGLE). Figure 3 also shows the location of the most massive FoF group (yellow stars) along with the next nine most massive FoF groups (white stars) in each protocluster. These are located in overdense regions where, in general, the gas temperatures are higher as a result of shock heating and feedback from stellar winds and/or black hole accretion. It is also apparent that regions with high H I fractions trace the overdense gas; photo-ionisation equilibrium with the UV background means the H I number density, n_{HI} , scales with the square of the gas density.

Finally, in Figure 4 we show the volume weighted distribution of the gas density, temperature and H I fraction for the protoclusters compared to the “field” (i.e. regions outwith R_{95}) in each of the simulations. The differences in gas properties between the protoclusters and the field are small, but as might be anticipated from an inspection of Figure 3, the protoclusters are slightly more dense, hotter and have a larger H I fraction. Figure 4 also highlights the differences in gas properties between the three different simulations. Although the gas density distribution is similar

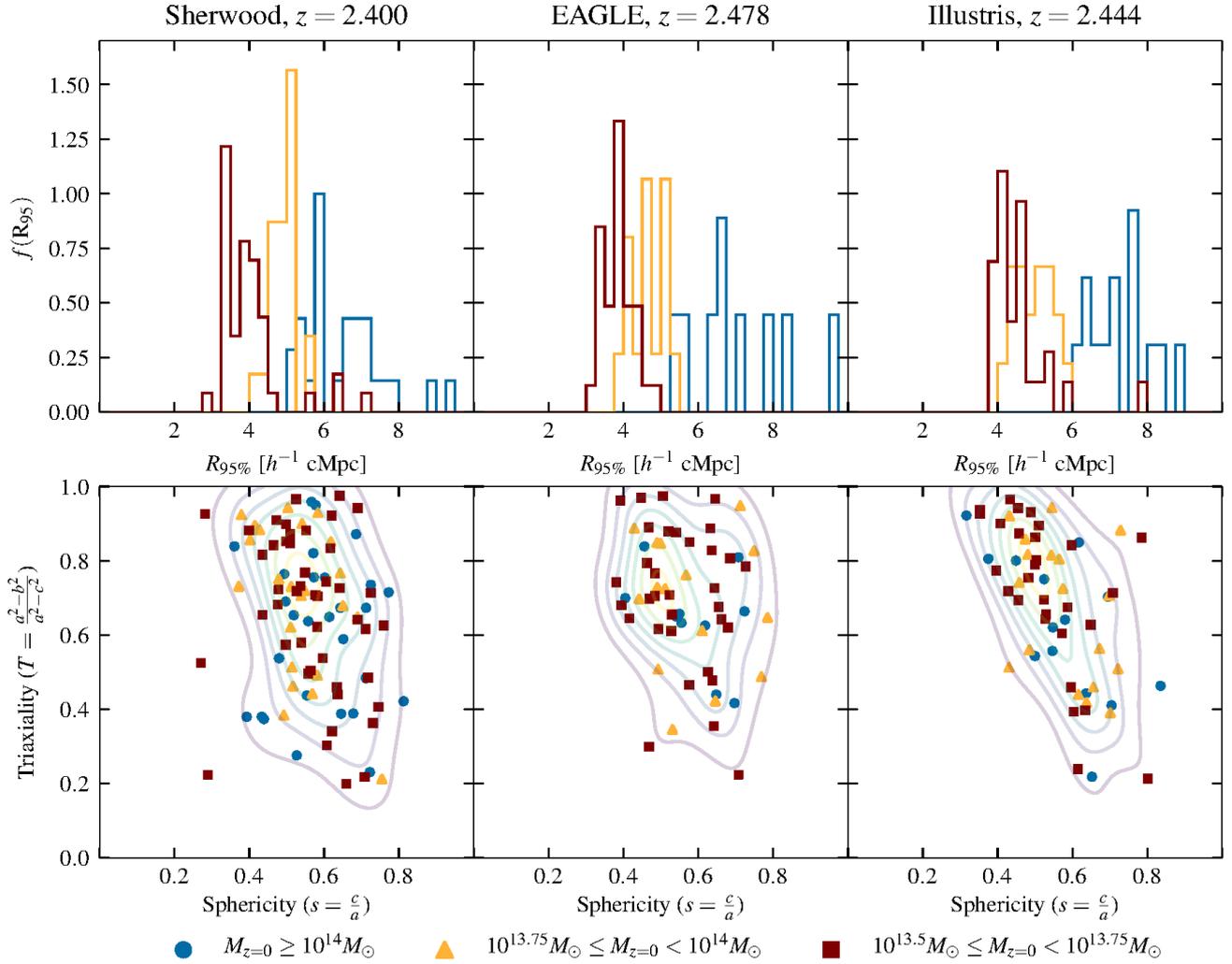


Figure 2. Top: Distribution of R_{95} for the protoclusters (blue) and large and small protogroups (orange and brown) in Sherwood (left), EAGLE (centre) and Illustris (right) at $z \sim 2.4$. Bottom: Scatter plots showing the triaxiality (where $T = 1$ and $T = 0$ are prolate and oblate spheroids, respectively) against the sphericity (where $s = 1$ and $s = 0$ are spherical and aspherical, respectively) of the protoclusters and protogroups, where a , b and c are the principal semi-axes and $a \geq b \geq c$. Contours show the kernel density estimate for all points.

in all three cases, there are differences in the temperature and H I fraction distribution that arise from the different UV background models (see Section 2). Illustris (EAGLE) has a slightly larger (smaller) average H I fraction compared to Sherwood. This is due to the smaller (larger) H I photo-ionisation rate used in the UV background model and – to a lesser extent – the dependence of the H I fraction on the lower (higher) gas temperature through the H II recombination rate.

5 CHARACTERISING MASS OVERDENSITIES IN Ly- α ABSORPTION ON $15h^{-1}$ cMpc SCALES

Observationally, protoclusters are identified as large scale, high density regions. For this reason, we first investigate the Ly- α absorption properties of mass overdensities within the hydrodynamical simulations, with no consideration as to whether these eventually collapse to form a cluster by

$z = 0$. From this, we determine the effectiveness of using Ly- α absorption to detect large scale overdense regions at $z = 2.4$. It is important to note, however, that this does not address how effectively Ly- α absorption probes the gas at $z \simeq 2.4$ that *actually* collapses to form clusters by $z = 0$ – we consider this further in Section 6.

5.1 Relationship between mass and effective optical depth on $15h^{-1}$ cMpc scales

In order to assess how well Ly- α absorption traces large scale, high density regions, we consider the correlation between mass and Ly- α effective optical depth on $15h^{-1}$ cMpc scales. We choose this scale as it corresponds to the characteristic size of protoclusters at $z \geq 2$ (Chiang et al. 2013; Muldrew et al. 2015), and also follows the scale adopted by Cai et al. (2016).

Our procedure throughout this work is as follows. First, we sum the masses of all particles in $15^3 h^{-3}$ cMpc³ cubic

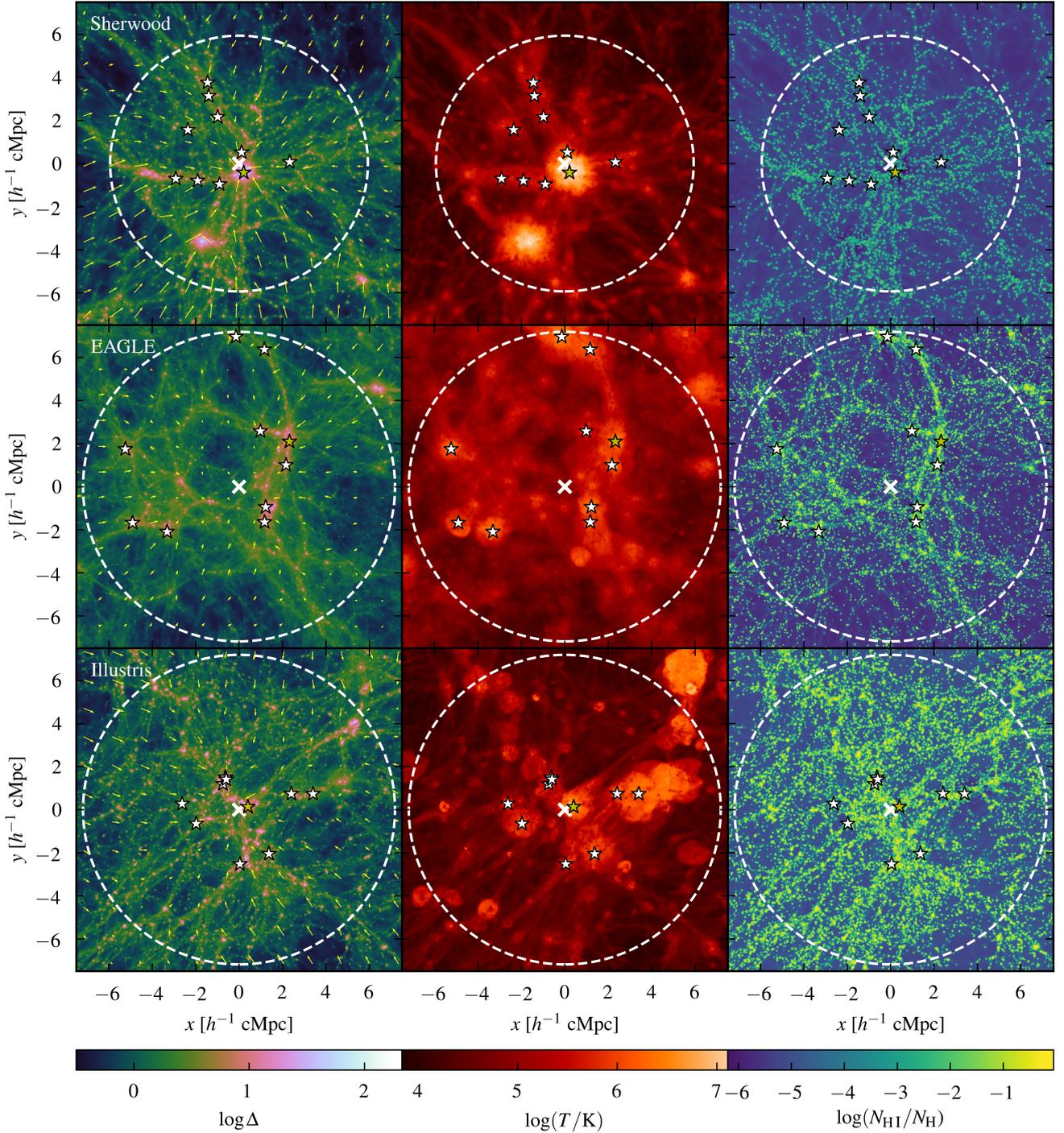


Figure 3. Projected maps of the normalised gas density, temperature and HI fraction for protoclusters at $z \sim 2.4$ with masses $M_{z=0} = 10^{14.3} M_{\odot}$, $M_{z=0} = 10^{14.4} M_{\odot}$ and $M_{z=0} = 10^{14.3} M_{\odot}$ in Sherwood (top row), EAGLE (middle row) and Illustris (bottom row). The projection depth along the z -axis is $15 h^{-1} \text{ cMpc}$, and is centred on the centre of mass of each protocluster. The dashed circles show R_{95} , and the star symbols correspond to the locations of the 10 most massive FoF groups within R_{95} . The yellow stars correspond to the most massive FoF groups in each protocluster, which have mass $M = 10^{13.6} M_{\odot}$, $M = 10^{12.8} M_{\odot}$ and $M = 10^{13.1} M_{\odot}$ in Sherwood, EAGLE and Illustris, respectively. The yellow arrows overlaid on the left column show the peculiar velocity field in the protoclusters relative to the centre of mass.

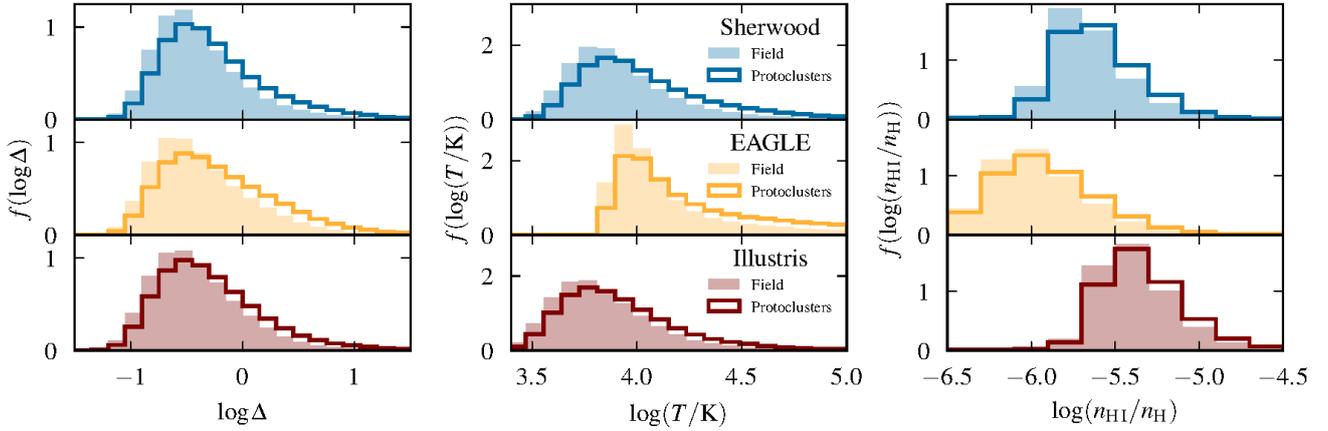


Figure 4. Volume weighted distributions of the normalised gas density, temperature and HI fraction in all protoclusters (line histograms) compared to the field (filled histograms) in Sherwood (blue), EAGLE (orange) and Illustris (brown).

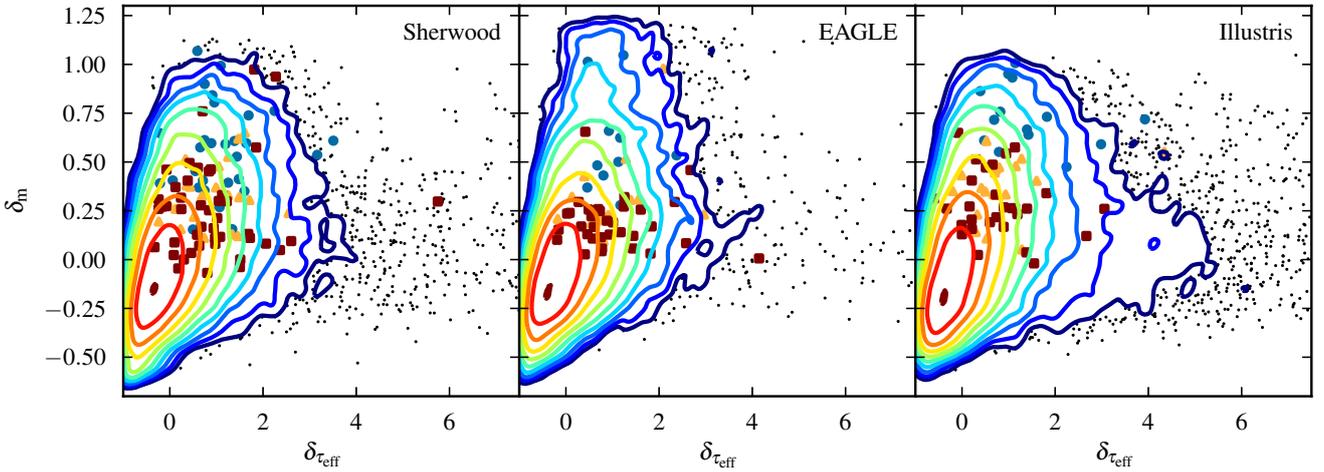


Figure 5. The relationship between δ_m and $\delta_{\tau_{\text{eff}}}$ on $15 h^{-1} \text{cMpc}$ scales (following Cai et al. 2016). Left to right, the panels show the results from Sherwood, EAGLE and Illustris. The contours represent the number density of data points relative to the central contour, with the number density decreasing by $1/3$ dex with each contour. Black points represent every system which lies outside of these contours (i.e. with relative number density $< 10^{-3}$). Additional points corresponding to the centre of mass of the protoclusters (blue circles) large protogroups (orange triangles) and small protogroups (brown squares) are also displayed.

volumes along every simulated Ly- α forest sight line (i.e. 30,000 in total for Sherwood, each with length $80 h^{-1} \text{cMpc}$), and compute the mass overdensity, $\delta_m = \frac{m - \langle m \rangle}{\langle m \rangle}$ in each volume. Hence, δ_m is a 3D average, where m is the total mass within each $15^3 h^{-3} \text{cMpc}^3$ volume, and $\langle m \rangle$ is the mass contained in a $15^3 h^{-3} \text{cMpc}^3$ volume with density equal to the mean density, $\langle \rho \rangle = \rho_{\text{crit}} \Omega_m (1+z)^3$, where

$$\langle m \rangle = 4.25 \times 10^{14} M_{\odot} \left(\frac{h}{0.678} \right)^{-1} \left(\frac{\Omega_m}{0.308} \right). \quad (3)$$

Masses above (below) this threshold represent overdense (underdense) volumes on $15 h^{-1} \text{cMpc}$ scales. Next, we associate every δ_m with the Ly- α forest effective optical depth, $\tau_{\text{eff}} = -\ln \langle F \rangle$, obtained from the $15 h^{-1} \text{cMpc}$ segments of simulated spectrum that pass directly through the centre

of each $15^3 h^{-3} \text{cMpc}^3$ volume⁶. We then compute $\delta_{\tau_{\text{eff}}} = \frac{\tau_{\text{eff}} - \langle \tau_{\text{eff}} \rangle}{\langle \tau_{\text{eff}} \rangle}$, where τ_{eff} is the effective optical depth measured in each $15 h^{-1} \text{cMpc}$ spectral segment, and $\langle \tau_{\text{eff}} \rangle = 0.20$ is obtained from Becker et al. (2013) at $z = 2.4$. In this way, we associate $\delta_{\tau_{\text{eff}}}$ from every 1D $15 h^{-1} \text{cMpc}$ segment with δ_m in the surrounding 3D $15^3 h^{-3} \text{cMpc}^3$ volume (see also Cai et al. 2016).

The relationship between δ_m and $\delta_{\tau_{\text{eff}}}$, averaged over $15 h^{-1} \text{cMpc}$ scales, is displayed in Figure 5. In Sherwood, EAGLE and Illustris the bulk of the simulation box corresponds to volumes close to the mean density, and there is a weak positive correlation between δ_m and $\delta_{\tau_{\text{eff}}}$. The centre of mass for all protoclusters and most protogroups (shown by the blue, orange and brown symbols) reside in overdense vol-

⁶ That is, the average transmission for each 1D spectral segment, $\langle F \rangle = e^{-\tau_{\text{eff}}}$, is obtained by averaging over all the pixels in a section of simulated spectrum with length $15 h^{-1} \text{cMpc}$.

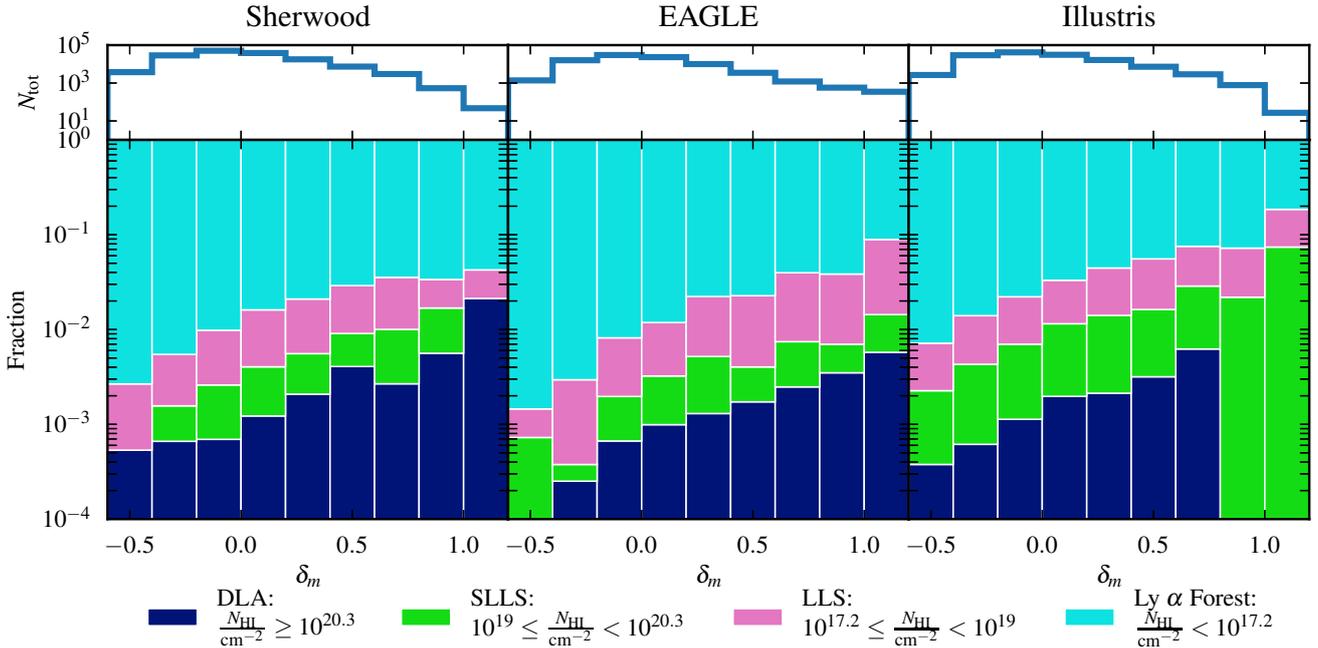


Figure 6. The fraction of $15h^{-1}$ cMpc segments in bins of $\Delta\delta_m = 0.2$ whose largest constituent H I column density measured on 50 km s^{-1} scales corresponds to either Ly- α forest, LLSs, SLLSs or DLAs in Sherwood (left), EAGLE (centre) and Illustris (right). The upper panels display the total number of $15h^{-1}$ cMpc segments in each bin.

umes, although note these are not usually associated with segments of high Ly- α opacity; the protocluster centre of mass will not always coincide with a halo or large value of τ_{eff} (cf. Figure 3). The differences between the three simulations are minimal at $\delta_{\tau_{\text{eff}}} \lesssim 2$, suggesting that variations in numerical methodology have little impact on the properties of the low density gas on large scales. By contrast, more significant differences between the simulations are apparent at $\delta_{\tau_{\text{eff}}} > 2$, corresponding to segments containing high column density, self-shielded absorption systems. These systems reside in both overdense and underdense volumes. The number of $\delta_{\tau_{\text{eff}}} > 2$ segments is greatest in Illustris, which is the simulation with the largest number of absorption systems at $10^{17} \leq N_{\text{HI}}/\text{cm}^{-2} \leq 10^{20}$ (see Figure 1).

Halos and galaxies are biased tracers of mass overdensity, therefore we expect a correlation between the presence of the high column density systems – responsible for the high $\delta_{\tau_{\text{eff}}}$ tail in Figure 5 – and δ_m . In Figure 6 the fraction of $15h^{-1}$ cMpc segments containing an absorption system, classed as either a DLA, SLLS, LLS or as Ly- α forest based on the largest constituent column density in each segment, is shown in bins of δ_m for each of the simulations. In all three simulations, the vast majority of the volume – irrespective of overdensity – is traced by Ly- α forest absorption, with 97.1, 98.0 and 93.5 per cent of $15h^{-1}$ cMpc segments containing systems with a maximum column density of $N_{\text{HI}} < 10^{17.2} \text{ cm}^{-2}$ in Sherwood, EAGLE and Illustris, respectively. At all δ_m , Illustris has a greater fraction of segments containing SLLSs than the other simulations, with ~ 20 per cent of all segments associated with volumes of $\delta_m > 1$ containing a LLS or SLLS. Note, however, that DLAs in Illustris are only present in segments associated with volumes of $\delta_m \leq 0.8$, implying that absorption systems in the most overdense volumes in Illustris have slightly

lower H I fractions relative to EAGLE and Sherwood. This is broadly consistent with the H I column density distribution, where Illustris contains an excess of absorption systems with $10^{17} \leq N_{\text{HI}}/\text{cm}^{-2} \leq 10^{20}$ relative to Sherwood and EAGLE. Figure 6 also demonstrates that damped systems with $N_{\text{HI}} > 10^{19} \text{ cm}^{-2}$ are present at all overdensities on $15h^{-1}$ cMpc scales, explaining the scatter with δ_m in the high $\delta_{\tau_{\text{eff}}}$ tail observed in Figure 5.

This highlights the importance of self-consistently modelling the incidence of high column density H I absorption systems when identifying overdense volumes using Ly- α absorption. Damped systems produce large values of $\delta_{\tau_{\text{eff}}}$ regardless of large-scale environment, and so segments of high $\delta_{\tau_{\text{eff}}}$ will not uniquely probe overdense volumes. In addition to this, in all three simulations there is an increase in the fraction of segments containing systems with $N_{\text{HI}} \geq 10^{19} \text{ cm}^{-2}$ with increasing δ_m . This implies that highly overdense volumes, such as those that may collapse to form a cluster by $z = 0$, will contain a greater fraction of sight lines that pass through a damped system.

5.2 CoSLAs: are they mass overdensities?

From Figure 6, it is evident that the majority of overdense volumes are traced by absorption from the Ly- α forest and LLSs. We now explore whether any of these overdense volumes would be classified as being associated with Coherently Strong intergalactic Ly- α Absorption systems (CoSLAs) within the hydrodynamical simulations. Cai et al. (2016) (hereafter C16) associate protoclusters with CoSLAs, defining these as all segments of Ly- α absorption with $\delta_{\tau_{\text{eff}}} > 3.5$ on $15h^{-1}$ cMpc scales after excluding any high column density absorbers with damping wings. Therefore, before se-

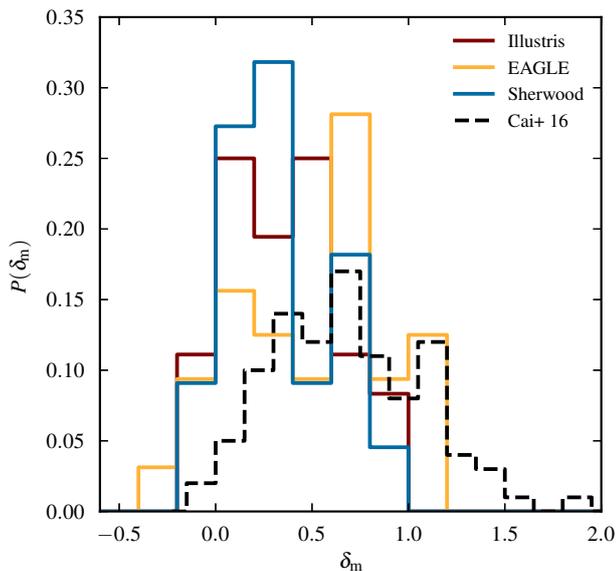


Figure 7. Probability distribution of δ_m for $15h^{-1}$ cMpc volumes with an associated $\delta_{\tau_{\text{eff}}} > 3.5$ (i.e. CoSLAs) after removing all sight lines containing damped systems in Sherwood (blue), EAGLE (orange) and Illustris (brown). The corresponding distribution from the analysis of a $1h^{-3}$ cGpc³ post-processed collisionless dark matter simulation by Cai et al. (2016) is displayed as a black dashed line. The total number of CoSLAs identified in each simulation is: Illustris (36), EAGLE (32), Sherwood (22).

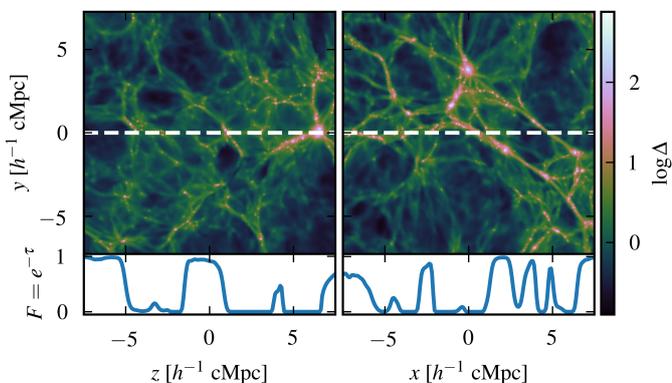


Figure 8. Examples of two $15h^{-1}$ cMpc volumes with an associated $\delta_{\tau_{\text{eff}}} > 3.5$ (i.e. CoSLAs) in Sherwood, one of which is drawn from an underdense volume with $\delta_m = -0.09$ (left), with the other corresponding to an overdense volume with $\delta_m = 0.98$ (right). In each case, the upper panels display the normalised gas density projected over a distance of $1h^{-1}$ cMpc centred on the sight line along which the Ly- α absorption spectrum is extracted (white dashed line), while the lower panels show the corresponding simulated Ly- α absorption. The underdense CoSLA has $\delta_{\tau_{\text{eff}}} = 3.76$ and a maximum $N_{\text{HI}} = 10^{18}$ cm⁻², while the overdense CoSLA has $\delta_{\tau_{\text{eff}}} = 3.68$ and a maximum $N_{\text{HI}} = 10^{15.3}$ cm⁻².

lecting CoSLAs, we first remove all sight lines that contain column densities $N_{\text{HI}} > 10^{19}$ cm⁻² in the simulated spectra. We take this approach to ensure that we not only remove the segments contain the damped system itself, but also any neighbouring segments where extended damping wings from these systems are still present. Due to the correlation between δ_m and the fraction of damped systems (see Figure 6),

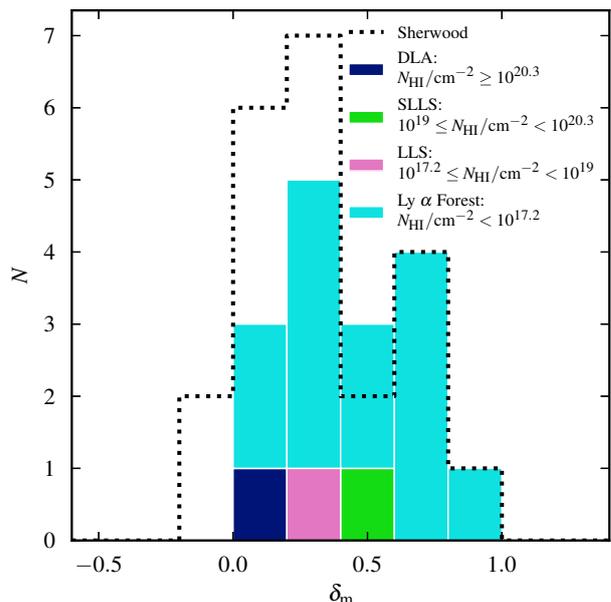


Figure 9. The filled histogram displays the number of $15h^{-1}$ cMpc volumes associated with segments of $\delta_{\tau_{\text{eff}}} > 3.5$ (i.e. CoSLAs) in bins of $\Delta\delta_m = 0.2$ for the QLy- α simulation. The colour of each stacked bar indicates the class of absorber each segment corresponds to in Sherwood, which is identical to QLy- α except for the lack of any cold ($T < 10^5$ K), dense ($\Delta > 10^3$) gas. The corresponding distribution for Sherwood is shown as the black dotted line, where there is a greater frequency of CoSLAs with $\delta_m < 0.4$. The total number of $15h^{-1}$ cMpc segments with $\delta_{\tau_{\text{eff}}} > 3.5$ in each simulation is: Sherwood (22), QLy- α (16).

this process preferentially removes segments corresponding to overdense volumes.

C16 used a collisionless dark matter simulation coupled with LyMAS (Peirani et al. 2014) to show that almost all CoSLAs correspond to mass overdensities in the range $0 < \delta_m < 2$, with ~ 70 per cent having $\delta_m \geq 0.5$. Chiang et al. (2013) showed that, at similar redshifts and scales ($z = 2$ and $16.3h^{-1}$ cMpc respectively), more than 80 per cent of volumes with $\delta_m \geq 0.8$ will collapse to form clusters with $M_{z=0} > 10^{14} M_{\odot}$ by $z = 0$. The expectation is therefore that most CoSLAs will also probe structures that collapse to form clusters by $z = 0$.

In Figure 7, we examine the δ_m associated with CoSLAs within each of the three hydrodynamical simulations. The distributions for Sherwood and Illustris are consistent with one another, with a median $\delta_m = 0.36 \pm 0.07$ and $\delta_m = 0.35 \pm 0.08$, respectively, where we estimate the uncertainty by bootstrapping with replacement. EAGLE has a higher median $\delta_m = 0.59 \pm 0.11$, formally differing from Illustris and Sherwood by 1.8σ . However, given the relatively small number of sight lines with coherent absorption on $15h^{-1}$ cMpc scales in the simulations, this may reflect differences in the rare, massive structures within each simulation. Regardless of the median of each distribution, however, in all three simulations the CoSLAs cover a broad δ_m range including $\delta_m \leq 0$, indicating that they do not exclusively probe large scale mass overdensities at $z = 2.4$. As a further illustration, Figure 8 displays examples of CoSLAs from both underdense ($\delta_m = -0.09$) and overdense ($\delta_m = 0.98$) volumes in

Sherwood. In both cases, we observe that a CoSLA arises from the alignment of structure along the line of sight. This can be due to LLSs associated with the alignment of several haloes punctuated by voids (left panel), or multiple lower column density absorbers associated with a more extended, filamentary structure (right panel).

These results differ from C16, who found CoSLAs to have a median $\delta_m = 0.75 \pm 0.03$ and almost exclusively correspond to overdense volumes. To explore the reason for this difference, we also consider the incidence of CoSLAs in the QLy- α simulation. The QLy- α model does not produce any damped systems and under produces LLSs (see Figure 1) due to missing high density gas, but is otherwise identical to Sherwood. Consequently, the QLy- α model should better approximate the lower resolution simulation⁷ used in C16.

In Figure 9, the δ_m distribution of CoSLAs in Sherwood is compared to the QLy- α simulation (this time in terms of number in each bin). The CoSLAs identified in QLy- α are coloured according to the maximum column density drawn from the matching locations within the Sherwood simulation. There are several points to note from Figure 9. First, QLy- α contains two CoSLAs that are in segments containing SLLSs or DLAs in Sherwood (meaning these segments were discarded), while Sherwood contains eight segments classified as CoSLAs that are not present in QLy- α . These eight CoSLAs all contain LLSs or high column density Ly- α forest systems in Sherwood, but have lower column densities – and therefore lower values of $\delta_{\tau_{\text{eff}}}$ – in QLy- α . Furthermore, all these additional systems have $\delta_m < 0.4$, which acts to dilute the correlation between δ_m and $\delta_{\tau_{\text{eff}}}$. This emphasises the importance of correctly capturing LLSs in the models; not only will a fraction of the CoSLAs in the QLy- α model be contaminated by the presence of damped systems, but additional systems with $10^{16} \lesssim N_{\text{HI}}/\text{cm}^{-2} \leq 10^{19}$ are missed, resulting in an erroneously strong correlation between δ_m and $\delta_{\tau_{\text{eff}}}$ for CoSLAs. Second, the CoSLAs in QLy- α have a median $\delta_m = 0.40 \pm 0.02$, with ~ 50 per cent of CoSLAs having $\delta_m \geq 0.5$. This median δ_m is consistent with the Sherwood simulation, although note it is still smaller when compared with C16.

5.3 Effect of box size and mass resolution on CoSLAs

Another important factor to consider is the box size (and therefore maximum cluster mass) and mass resolution used in the simulations. We examine how mass resolution and box size impact on the relationship between mass overdensities and $\delta_{\tau_{\text{eff}}}$ on $15h^{-1}$ cMpc scales using four QLy- α runs drawn from the Sherwood simulation suite. These simulations are summarised in Table 2. As previously, sight lines were extracted with the same average transverse separation of $0.8h^{-1}$ cMpc from all runs.

The effect of an increase in box size on the δ_m and $\delta_{\tau_{\text{eff}}}$ distributions for CoSLAs is displayed in Figure 10. The δ_m

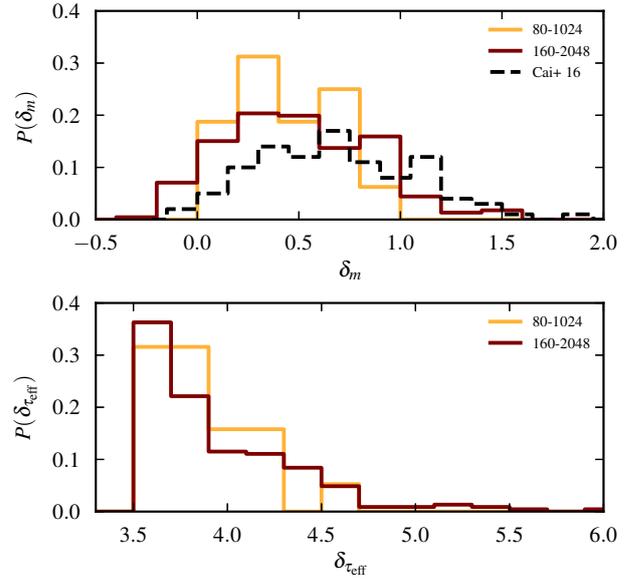


Figure 10. The probability distribution of CoSLAs as a function of δ_m (upper panel) and $\delta_{\tau_{\text{eff}}}$ (lower panel) for the QLy- α model (80-1024) compared to a simulation with the same mass resolution but a volume eight times larger (160-2048). The δ_m distribution from C16 is displayed as the dashed black histogram in the upper panel.

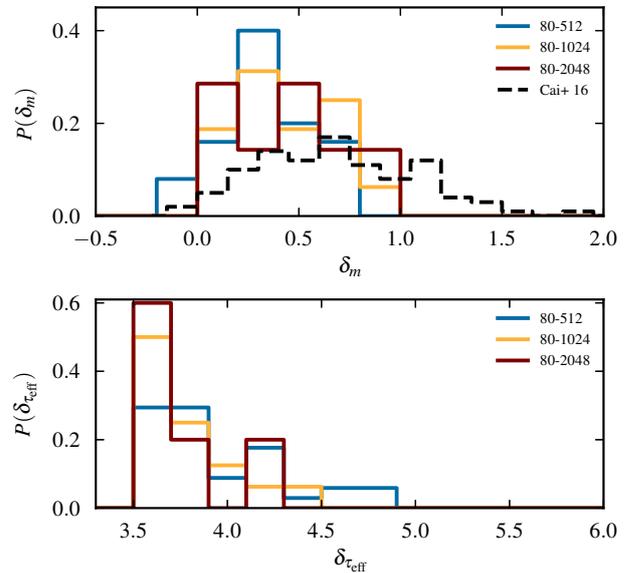


Figure 11. The probability distribution of CoSLAs as a function of δ_m (upper panel) and $\delta_{\tau_{\text{eff}}}$ (lower panel) for the QLy- α model (80-1024) compared to simulations with the same box size but a particle mass that is eight times larger (80-512) or smaller (80-2048). The δ_m distribution from C16 is displayed as the dashed black histogram in the upper panel.

distribution in the largest volume simulation (160-2048) extends toward higher values as a consequence of the rarer, more massive systems present in the larger volume. The number of CoSLAs per unit volume increases by 77 per cent in the 160-2048 simulation (see Table 2) relative to the fiducial 80-1024 (QLy- α) model, although both the 80-

⁷ C16 use a collisionless dark matter simulation in a $1h^{-3}$ cGpc³ volume with 1024^3 particles, yielding a particle mass of around $9.6 \times 10^{10} M_{\odot}$. For comparison, the typical dark matter particle mass needed to resolve the small scale structure of the Ly- α forest at $z \simeq 2$ is $\sim 10^7 M_{\odot}$ (Bolton & Becker 2009).

Table 2. Hydrodynamical simulations used to test the effect of box size and mass resolution on CoSLAs. All four simulations were performed using the quick-Ly- α approach that excludes dense, star forming gas (see Section 2). The columns list, from left to right: the simulation name, box size, number of resolution elements, dark matter and gas particle masses, and the number of CoSLAs per unit volume obtained using sight lines with an average transverse separation of $0.8h^{-1}$ cMpc. All of the $80h^{-1}$ cMpc boxes were performed with initial conditions generated using the same random seed.

Name	Box size [h^{-1} cMpc]	N_{tot}	M_{dm} [M_{\odot}]	M_{gas} [M_{\odot}]	N_{CoSLAs}/V [h^3 cMpc $^{-3}$]
80-512	80	2×512^3	4.06×10^8	7.52×10^7	4.9×10^{-5}
80-1024	80	2×1024^3	5.07×10^7	9.41×10^6	3.1×10^{-5}
80-2048	80	2×2048^3	6.34×10^6	1.17×10^6	1.4×10^{-5}
160-2048	160	2×2048^3	5.07×10^7	9.41×10^6	5.5×10^{-5}

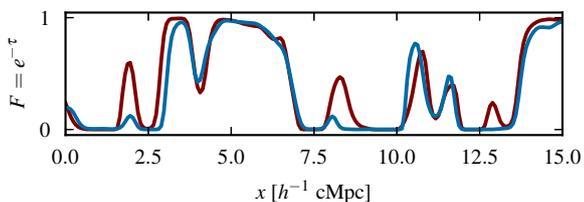


Figure 12. Simulated CoSLAs from the same segment drawn from the 80-512 (blue) and 80-2048 (dark red) models. This segment corresponds to $\delta_{\tau_{\text{eff}}} = 3.90$ in 80-512 but only $\delta_{\tau_{\text{eff}}} = 3.67$ in 80-2048, due to the increased transmission from the better resolved underdense gas in the higher resolution model.

1024 and 160-2048 models share similar median values of $\delta_m = 0.40 \pm 0.02$ and $\delta_m = 0.48 \pm 0.05$, respectively. The corresponding $\delta_{\tau_{\text{eff}}}$ distribution displays similar behaviour, with a tail that extends to larger values of $\delta_{\tau_{\text{eff}}}$ in the 160-2048 model. The effect of the ~ 1000 times larger volume used by C16 therefore likely explains some of the differences relative to this work. The CoSLAs selected in C16 (shown by the dashed black histogram in the upper panel of Figure 10) probe rarer, more massive systems that are not present in our small, higher resolution models. We are unable to reliably assess how effectively CoSLAs probe overdense $15^3 h^{-3}$ cMpc 3 volumes with $\delta_m > 1$ using our fiducial $80^3 h^{-3}$ cMpc $^{-3}$ simulation box. However, this still does not fully explain the greater incidence of CoSLAs at overdensities of $\delta_m < 0.5$ within the hydrodynamical simulations compared to C16, even when using QLy- α models that underpredict the number of CoSLAs at $\delta_m \leq 0.4$ due to missing LLSs. A possible explanation is that the lack of high τ_{eff} systems at δ_m is even more pronounced in much larger collisionless simulations, due to unresolved high column density absorption from small scale structure, but the exact reason for this difference remains unclear.

Mass resolution also has an important – but more subtle – effect on the typical $\delta_{\tau_{\text{eff}}}$ and δ_m associated with CoSLAs. The top panel of Figure 11 highlights how the δ_m distribution of CoSLAs is shifted toward lower values for the 80-512 simulation when compared to the two higher resolution models. The median $\delta_m = 0.38 \pm 0.01$ for the 80-512 model, compared to median values of $\delta_m = 0.40 \pm 0.02$ and $\delta_m = 0.42 \pm 0.04$ for 80-1024 (QLy- α) and 80-2048, respectively. However, the average number of CoSLAs per unit volume also decreases by a factor ~ 2 when increasing the mass resolution by a factor of 8 (see Table 2), suggesting this quantity is not yet converged with mass resolution.

The lower panel of Figure 11 elucidates the origin of this behaviour: at higher mass resolution the high $\delta_{\tau_{\text{eff}}}$ tail of the CoSLA distribution is truncated in comparison with the lower resolution runs. This is a consequence of the failure of the lower resolution runs to correctly resolve the structure of the Ly- α absorption, particularly in underdense regions, which are too opaque in low mass resolution models (Bolton & Becker 2009). The effect of this poorly resolved underdense gas is to increase the number of volumes on $15h^{-1}$ cMpc scales that are associated with segments which exceed the CoSLA identification threshold of $\delta_{\tau_{\text{eff}}} = 3.5$. An illustration of this effect is shown in Figure 12, where Ly- α absorption from a randomly selected segment in the 80-2048 simulation (dark red) is compared directly to the 80-512 simulation (blue). This implies that – in addition to not capturing absorption from high column density, self-shielded absorbers – low mass resolution models that do not adequately resolve the structure of the Ly- α forest on small scales will overpredict the incidence of CoSLAs above a fixed $\delta_{\tau_{\text{eff}}}$ threshold.

6 CHARACTERISING PROTOCLUSTERS IN Ly- α ABSORPTION

We now address the question of how protoclusters, as opposed to overdensities, are characterised by Ly- α absorption. Whilst – observationally – protoclusters are identified as overdense regions at $z \sim 2$, when using simulations we have access to *a priori* knowledge of which structures will collapse to form a cluster along with the associated cluster mass. We now use this simulation based “true” definition of a protocluster to examine the opacity of protoclusters at $z = 2.4$ on $15h^{-1}$ cMpc scales. We do not expect variations in the numerical methodology used by Sherwood, EAGLE and Illustris to significantly change these results, so from this point onward we focus on analysing only the Sherwood simulation.

In Figure 13, the fraction of all $15h^{-1}$ cMpc segments in bins of $\Delta\delta_{\tau_{\text{eff}}} = 0.3$ that are associated⁸ with protoclusters (blue shading) or protogroups (orange and brown shading) are displayed in the central panel. In the event that a system is associated with more than one protocluster, it is

⁸ A $15h^{-1}$ cMpc segment is associated with a protocluster or protogroup if at least one third of the segment passes within R_{95} of any protocluster/group.

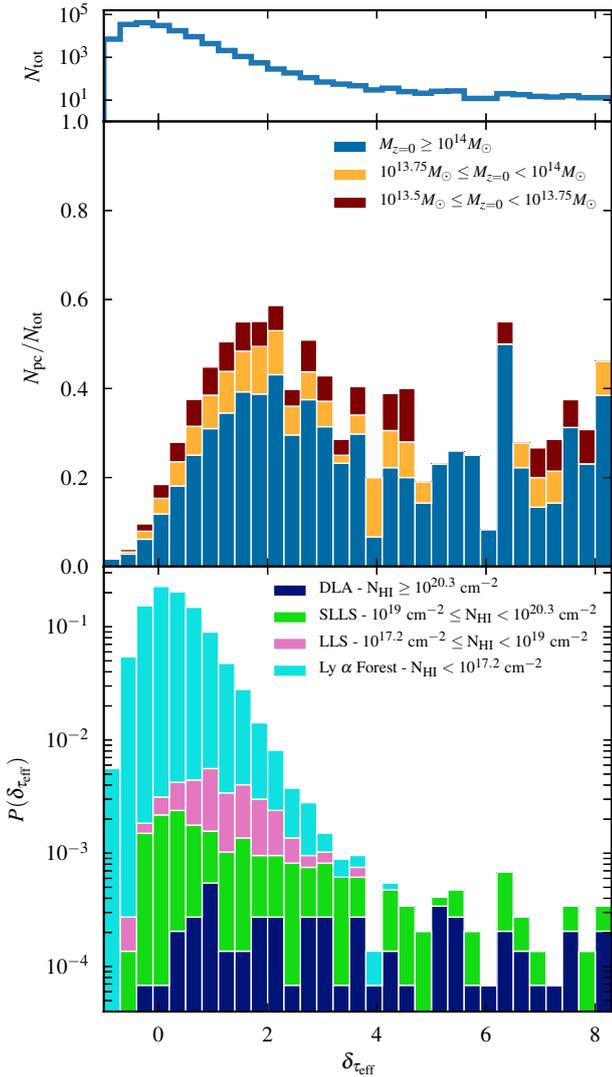


Figure 13. Top: Total number of $15h^{-1}$ cMpc segments in bins of $\Delta\delta_{\tau_{\text{eff}}} = 0.3$. Middle: The fraction of $15h^{-1}$ cMpc segments in bins of $\Delta\delta_{\tau_{\text{eff}}} = 0.3$ associated with protoclusters or protogroups. The colour of each stacked bar indicates the mass of the protocluster or protogroup associated with the absorption. Bottom: probability distribution for all $15h^{-1}$ cMpc segments that are associated with protoclusters. The colour of each stacked bar indicates the maximum column density within each segment.

associated with the most massive one only. The corresponding number of $15h^{-1}$ cMpc segments in each bin is shown in the upper panel, with most segments close to the mean, $\delta_{\tau_{\text{eff}}} = 0$.

Note that in almost all bins, the majority of the “associated” segments are in the protoclusters with $M_{z=0} \geq 10^{14} M_{\odot}$. This in part reflects the fact that the protoclusters occupy a larger fraction of the simulation volume relative to the protogroups⁹. Typically fewer than 10 per cent of segments with $\delta_{\tau_{\text{eff}}} \lesssim 0$ are associated with protoclusters, increasing to ~ 40 per cent at $\delta_{\tau_{\text{eff}}} = 2$. At $\delta_{\tau_{\text{eff}}} \gtrsim 2$, how-

⁹ Protoclusters occupy 7.7 per cent of the Sherwood simulation volume at $z = 2.4$, whereas the large and small protogroups together only occupy 5.0 per cent.)

ever, this fraction declines, where on average only ~ 25 per cent of segments are associated with protoclusters. Figure 13 therefore demonstrates that for $\delta_{\tau_{\text{eff}}} < 2$ there is a positive, albeit weak, correlation between $\delta_{\tau_{\text{eff}}}$ and the likelihood of the associated volume collapsing to form a cluster by $z = 0$.

In the lower panel of Fig. 13, the probability distribution for $15h^{-1}$ cMpc segments residing within the protoclusters is displayed as a function of $\delta_{\tau_{\text{eff}}}$, with the shading indicating the fraction of segments in each bin that contains a given maximum H I column density. The majority of segments passing through protoclusters only contain Ly- α forest absorption systems and have $\delta_{\tau_{\text{eff}}} \sim 0$: 28 per cent of the segments which are associated with protoclusters are Ly- α forest with $\delta_{\tau_{\text{eff}}} < 0$, and 84 per cent are entirely Ly- α forest with $\delta_{\tau_{\text{eff}}} < 1$. Thus, while the majority of segments with $\delta_{\tau_{\text{eff}}} < 1$ do not correspond to protoclusters, but rather to “field” regions, the majority of segments associated with protoclusters lie in the same $\delta_{\tau_{\text{eff}}}$ range and share identical characteristics. This reflects the fact that protoclusters at $z \simeq 2.4$ are extended over large scales (Muldrew et al. 2015); this means that it is very difficult – or impossible – to disentangle most protocluster sight lines from the field on $15h^{-1}$ cMpc scales.

At larger values of $\delta_{\tau_{\text{eff}}}$ the segments associated with protoclusters become dominated by SLLSs and DLAs. At $\delta_{\tau_{\text{eff}}} > 2.9$, the majority of segments associated with protocluster contain damped systems, and 96 per cent of the segments with $\delta_{\tau_{\text{eff}}} \gtrsim 3.5$ (i.e. the threshold for CoSLAs) that are associated with protoclusters arise from high density gas producing damped Ly- α absorption. At $\delta_{\tau_{\text{eff}}} > 3.5$ SLLSs and DLAs are the dominant probe of gas that forms clusters by $z = 0$, but – critically for protocluster selection by Ly- α absorption – the middle panel of Figure. 13 shows that these are not uniquely associated with protoclusters at $z \simeq 2.4$, but are more likely to be in the field.

Figure 14 displays selected examples of segments that pass through protoclusters that cover the full range of $\delta_{\tau_{\text{eff}}}$ observed in the simulation. Most sight lines passing through protoclusters with $\delta_{\tau_{\text{eff}}} < 2$ are exemplified by the two left-hand panels in Figure 14. In these cases the sight line passes through two filamentary structures perpendicular to the sight line (left panel, $\delta_{\tau_{\text{eff}}} = -0.20$) or through a small region of overdense gas within a protocluster (second left panel, $\delta_{\tau_{\text{eff}}} = 1.72$), but in both cases do not exhibit extended regions of saturated Ly- α absorption.

By contrast, in the two right-hand panels of Figure 14, two examples of segments that pass through protoclusters and have $\delta_{\tau_{\text{eff}}} > 3.5$ (i.e. the CoSLA threshold) are displayed. The right-most panel ($\delta_{\tau_{\text{eff}}} = 4.65$) displays strong absorption due to the presence of an extended damping wing arising from a DLA; there is no large scale gas overdensity present along the sight line. The segment displayed on the second right ($\delta_{\tau_{\text{eff}}} = 4.57$) has no damped absorbers, but instead the absorption arises from haloes and a filament that are aligned with the sight line and are associated with two intersecting protoclusters. The high value of $\delta_{\tau_{\text{eff}}}$ in this case results from extended Ly- α forest absorption. This implies that it is the orientation of dense gas with respect to the sight line that has the greatest impact on the value of $\delta_{\tau_{\text{eff}}}$, rather than either the overdensity δ_m (see also Figure 8) or the presence of a protocluster.

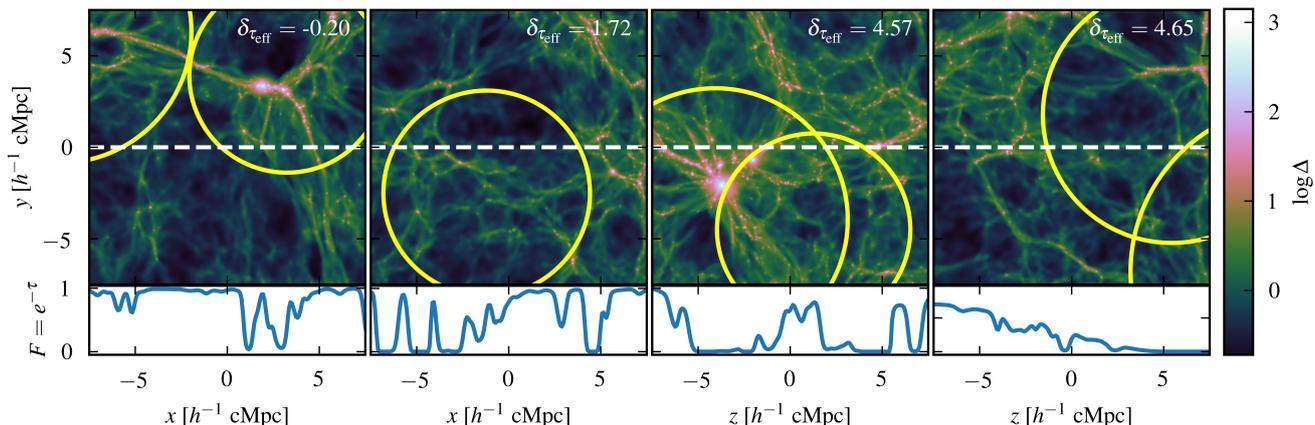


Figure 14. Example of four different $15h^{-1}$ cMpc segments that pass through protoclusters in Sherwood. The top panels show projected maps of the normalised gas density in a $15h^{-2}$ cMpc² slice with a projection depth of $1h^{-1}$ cMpc. The white dashed line shows the direction in which the Ly- α absorption spectrum was extracted, while the yellow circles indicate the cross section of all protoclusters that intersect with the slice. The bottom panels show the corresponding mock Ly- α absorption spectra. From left to right, each segment corresponds to: a low opacity segment with $\delta_{\tau_{\text{eff}}} = -0.20$ containing a maximum HI column density $N_{\text{HI}} = 10^{13.8}$ cm⁻², a higher opacity segment with $\delta_{\tau_{\text{eff}}} = 1.72$ and maximum $N_{\text{HI}} = 10^{14.8}$ cm⁻², a CoSLA with $\delta_{\tau_{\text{eff}}} = 4.57$ and maximum $N_{\text{HI}} = 10^{16.9}$ cm⁻², and a DLA with $\delta_{\tau_{\text{eff}}} = 4.65$ and maximum $N_{\text{HI}} = 10^{20.5}$ cm⁻².

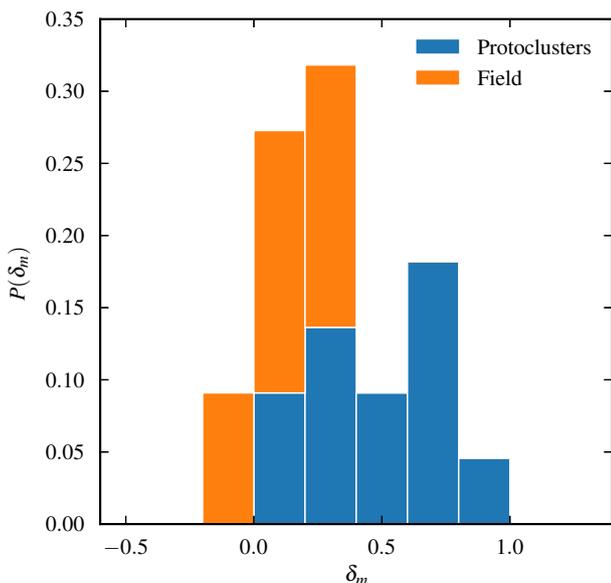


Figure 15. The probability distribution of CoSLAs identified in Sherwood as a function of δ_m , coloured by whether the CoSLA is associated with a protocluster (12; blue) or associated with the field (10; orange).

7 SELECTING PROTOCLUSTERS WITH LINE OF SIGHT Ly- α ABSORPTION

Now that we have characterised the Ly- α absorption associated with protoclusters, we finally turn to examine the effectiveness with which one can select protocluster regions using line of sight Ly- α absorption. We have established that most of the sight lines that pass through protoclusters with $M_{z=0} \sim 10^{14} M_{\odot}$ exhibit low values of $\delta_{\tau_{\text{eff}}}$, and it is impossible to distinguish these from the field. On the other hand, for large values of $\delta_{\tau_{\text{eff}}}$ most of the segments associated with protoclusters are the result of damped absorption systems

with $N_{\text{HI}} > 10^{19}$ cm⁻², but these systems do not uniquely trace protoclusters.

In addition, in the simulations there are a small subset of high opacity segments within protoclusters that do not exhibit damped absorption (e.g. the example CoSLA shown in the second from right panel in Figure 14). This is demonstrated further in Figure 15, which shows the distribution of CoSLAs in Sherwood split into two groups as a function of δ_m : those that are associated with a protocluster and those that are not. We find 55 per cent of CoSLAs are associated with protoclusters with a median $\delta_m = 0.4 \pm 0.1$, in comparison to CoSLAs that probe the field, with median $\delta_m = 0.06 \pm 0.05$. This indicates that, while the two populations cannot be separated by their Ly- α absorption spectra alone, approximately half of CoSLAs – which in general arise from the alignment of overdense structure along the line of sight – are indeed associated with protoclusters. Thus we find that CoSLAs are not a unique tracer of protoclusters, even when assuming perfect removal of damped systems from the mock data.

The effective optical depth threshold that defines a CoSLA as a segment exhibiting $\delta_{\tau_{\text{eff}}} > 3.5$ on $15h^{-1}$ cMpc scales is to some extent arbitrary. We may therefore consider whether a fixed threshold in $\delta_{\tau_{\text{eff}}}$ can be chosen that minimises the contamination within a sample of protoclusters selected from the mock spectra. We define the contamination as $N(\geq \delta_{\tau_{\text{eff}}})/N_{\text{tot}}(\geq \delta_{\tau_{\text{eff}}})$, where $N(\geq \delta_{\tau_{\text{eff}}})$ is the number of $15h^{-1}$ cMpc segments above the $\delta_{\tau_{\text{eff}}}$ threshold that do not pass through at least $5h^{-1}$ cMpc of a protocluster volume, and $N_{\text{tot}}(\geq \delta_{\tau_{\text{eff}}})$ is the total number of $15h^{-1}$ cMpc segments above the $\delta_{\tau_{\text{eff}}}$ threshold in the entire simulation volume. We also define the sample “completeness” for a given $\delta_{\tau_{\text{eff}}}$ threshold as the fraction of protoclusters with at least one sight line that exhibits absorption on $15h^{-1}$ cMpc scales above the $\delta_{\tau_{\text{eff}}}$ threshold. This is defined as $N_{\text{pc}}(\geq \delta_{\tau_{\text{eff}}})/N_{\text{pc,tot}}(\geq \delta_{\tau_{\text{eff}}})$, where $N_{\text{pc}}(\geq \delta_{\tau_{\text{eff}}})$ is the number of unique protoclusters probed by $15h^{-1}$ cMpc segments above

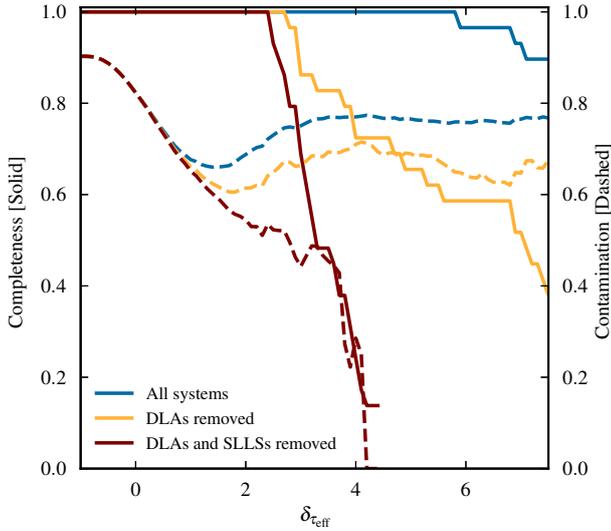


Figure 16. Contamination (dashed) and “completeness” (solid) for protoclusters (with $M_{z=0} \geq 10^{14} M_{\odot}$) when selecting all segments above a fixed threshold of $\delta\tau_{\text{eff}}$ on $15h^{-1}$ cMpc scales (see text for details). The blue curves correspond to the case where all absorption is considered, orange curves to the case where sight lines containing DLAs are removed, and brown curves to the case where SLLSs and DLAs are removed. Results are displayed only where there is more than one $15h^{-1}$ cMpc segment in each bin.

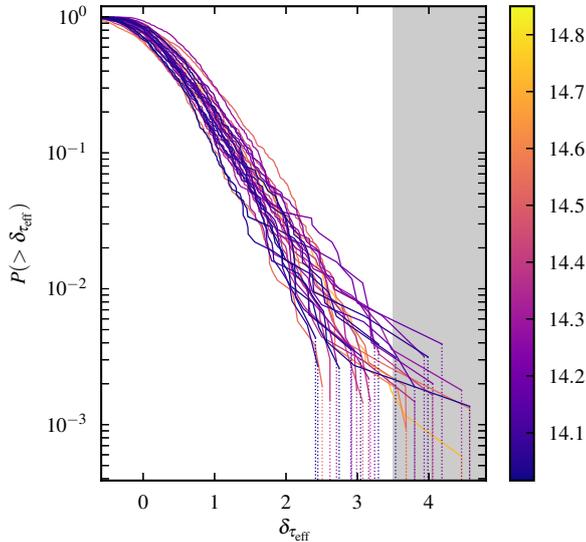


Figure 17. Reverse cumulative distribution functions, $P(> \delta\tau_{\text{eff}})$, for all $15h^{-1}$ cMpc segments that are associated with each of the 29 protoclusters in Sherwood. Each line is coloured according to the mass of the cluster at $z = 0$. The grey shaded region indicates the area above the CoSLA selection threshold of $\delta\tau_{\text{eff}} > 3.5$.

the $\delta\tau_{\text{eff}}$ threshold, and $N_{\text{pc,tot}}$ is the total number of protoclusters in the simulation (i.e. 29 for Sherwood, see Table 1).

These values are plotted as a function of the $\delta\tau_{\text{eff}}$ threshold in Figure 16 for three different cases: where absorption in all the sight lines is considered (blue curves), where sight lines containing DLAs are removed from the sample (orange curves), and the case where both SLLSs and DLAs are re-

moved (brown curves). When considering all sight lines, the protocluster sample remains 100 per cent complete using a threshold up to $\delta\tau_{\text{eff}} \sim 6$ because all protoclusters contain DLA systems that produce such high $\delta\tau_{\text{eff}}$. The contamination also remains very high at ~ 75 per cent, which reflects the fact that the damped Ly- α absorbers do not uniquely probe protocluster environments. If we instead assume perfect removal of all sight lines containing DLAs, the contamination rate falls slightly to around 65 per cent for the segments with the strongest absorption. Finally, removing all the $15h^{-1}$ cMpc segments containing SLLSs as well as DLAs causes the contamination to drop to zero at $\delta\tau_{\text{eff}} = 4.1$, i.e. all the remaining segments with this $\delta\tau_{\text{eff}}$ uniquely trace protocluster gas. Consequently, if we assume perfect removal of damped absorption systems, it is possible to obtain a perfectly clean sample of protoclusters by applying a sufficiently high threshold in $\delta\tau_{\text{eff}}$.

This clean sample of protoclusters is, however, highly incomplete as not all protoclusters will exhibit strong coherent Ly- α absorption on $15h^{-1}$ cMpc scales. This can be seen in Figure 16, where the drop in contamination is accompanied by the completeness falling to only 17 per cent. This means that less than a fifth¹⁰ of protoclusters have any sight lines with $\delta\tau_{\text{eff}} > 4.1$. This is further exemplified in Figure 17, which shows the reverse cumulative distribution function of $\delta\tau_{\text{eff}}$ for each of the 29 protoclusters in Sherwood. The colour of each line corresponds to the mass of the resulting cluster at $z = 0$. This demonstrates that high $\delta\tau_{\text{eff}}$ segments that pass through protoclusters are rare: less than 0.1 per cent of segments that pass through $M_{z=0} \sim 10^{14} M_{\odot}$ protoclusters have $\delta\tau_{\text{eff}} > 3.5$ – corresponding to the CoSLA threshold defined by C16, shown by the grey shading.

All the CoSLAs associated with protoclusters in Sherwood are listed in detail in Table 3: a total of 14 unique protoclusters over a mass range of $10^{14.0} - 10^{14.7} M_{\odot}$ are probed by 12 unique CoSLAs, and 42 per cent of the CoSLAs associated with protoclusters pass through more than one protocluster. A sample of four of the CoSLAs associated with protoclusters in Sherwood are displayed in Figure 18. We observe again that typically there is an alignment of structure along the sight line that causes the coherent Ly- α absorption. Orientation of structure to the line of sight, rather than association with a protocluster of a given mass or an overdensity, appears to be a critical factor that determines the extended nature of the Ly- α absorption. In this context, we briefly note that Finley et al. (2014) argued for the detection of an intergalactic filament based on observations of multiple LLSs and SLLSs with $N_{\text{HI}} > 10^{18} \text{ cm}^{-2}$ along two closely separated quasar sight lines at $z = 2.69$. The seven strong H I absorption systems observed by Finley et al. (2014) span $\sim 1700 \text{ km s}^{-1}$, corresponding to $16.6h^{-1}$ cMpc at $z = 2.69$. While Finley et al. (2014) could not definitively rule out association of the H I absorbers with a protocluster, their favoured interpretation is broadly consistent with our analysis. Taken a step further, this suggests that CoSLAs may in fact be a tracer of extended filamentary structure in the

¹⁰ The completeness is likely to be strongly dependent on the protocluster mass. Since we do not probe massive protoclusters – due to our small box size – this completeness is likely to be a lower limit.

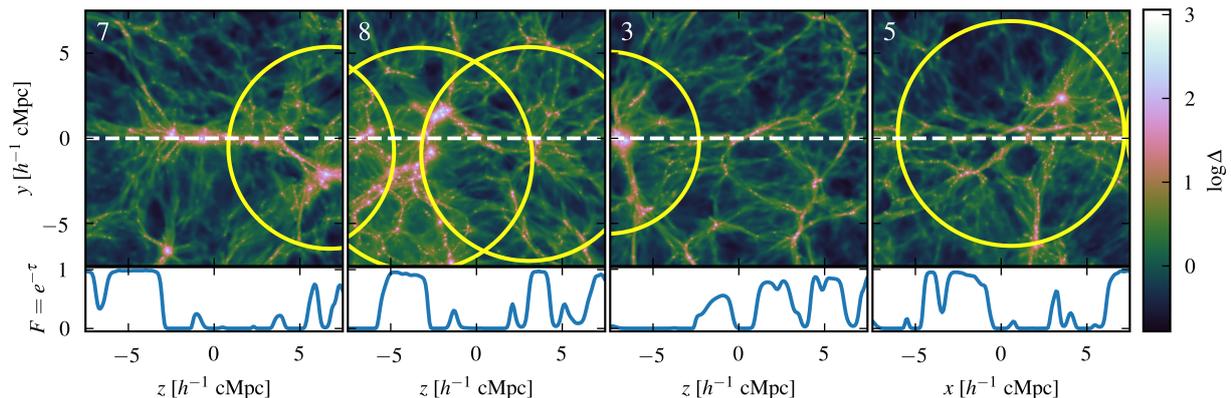


Figure 18. Example of four different CoSLAs that pass through protoclusters in Sherwood. The top panels show projected maps of the normalised gas density in a $15h^{-2}$ cMpc² slice with a projection depth of $1h^{-1}$ cMpc. The white dashed line shows the direction in which the Ly- α absorption spectrum was extracted. The bottom panels show the corresponding mock Ly- α absorption spectra. The numbers in the upper left of each panel correspond to the CoSLAs listed in Table 3. From left to right, the maximum H I column density associated with each CoSLA is $N_{\text{HI}} = 10^{16.5}$ cm⁻², $10^{14.9}$ cm⁻², $10^{19.0}$ cm⁻² and $10^{14.7}$ cm⁻².

Table 3. All CoSLAs (defined as $15h^{-1}$ cMpc segments with $\delta_{\tau_{\text{eff}}} > 3.5$ that do not contain a damped Ly- α absorber) in Sherwood that are associated with protoclusters. A total of 14 protoclusters are probed by 12 unique CoSLAs. Note that each CoSLA may probe multiple protoclusters.

CoSLA	$\delta_{\tau_{\text{eff}}}$	δ_m	Protocluster $M_{z=0}$ [$\log_{10}(M_{\odot})$]
1	4.57	0.56	14.54, 14.04
2	4.46	0.80	14.72, 14.28
3	4.18	0.38	14.19
4	4.05	0.37	14.25
5	4.05	0.42	14.41
6	3.99	0.12	14.05
7	3.93	0.35	14.22
8	3.80	0.72	14.35, 14.18
9	3.68	0.98	14.59, 14.53
10	3.67	0.70	14.72, 14.53
11	3.56	0.62	14.59
12	3.53	0.12	14.06

early Universe. An intriguing possibility is that CoSLAs, and their association (if any) with galaxies and/or metal absorption lines, may therefore provide a route to identifying and studying filamentary environments at $z > 2$.

8 CONCLUSIONS

In this work we have used state of the art hydrodynamical simulations from the Sherwood (Bolton et al. 2017), EAGLE (Schaye et al. 2015; Crain et al. 2015; McAlpine et al. 2016) and Illustris (Vogelsberger et al. 2014; Nelson et al. 2015) projects to examine the signature of protoclusters observed in Ly- α absorption at $z \simeq 2.4$. Building upon earlier work using low resolution collisionless dark matter simulations (e.g. Stark et al. 2015; Cai et al. 2016), here we use models that resolve small scale structure in the IGM, correctly reproduce the incidence of H I absorption systems (including high column densities that are self-shielded to Lyman continuum photons), and track the formation of structure to $z = 0$.

We examine the impact of small scale gas structure on the signature of large scale overdensities on $15h^{-1}$ cMpc scales, finding that the simulation mass resolution required for resolving Ly- α absorption on small scales is also a requirement for correctly modelling the average τ_{eff} on $15h^{-1}$ cMpc scales. This is necessary for capturing the incidence of H I absorption systems over a wide range of column densities, including damped systems and LLSs. At the same time, however, adequate mass resolution is necessary for correctly capturing the opacity of underdense regions in the IGM. A mass resolution that is too low will overpredict the typical Ly- α effective optical depth on $15h^{-1}$ cMpc scales.

We furthermore assess the prevalence of coherent Ly- α absorption within protoclusters at high redshift. Our main conclusions may be summarised as follows:

- We confirm there is a weak correlation between the mass overdensity, δ_m , and the effective optical depth relative to the mean, $\delta_{\tau_{\text{eff}}}$, on a $15h^{-1}$ cMpc scales, in the simulations, although there is a large amount of scatter that, particularly at large values of $\delta_{\tau_{\text{eff}}}$, means it is not possible to uniquely identify large scale overdensities with strong Ly- α absorption. This remains true even if first removing all damped Ly- α absorption systems that arise from dense, neutral gas on small scales.
- We examine the properties of coherently strong intergalactic Ly- α absorption systems (CoSLAs) in the simulations. CoSLAs – defined by C16 as regions on $15h^{-1}$ cMpc scales with 4.5 times the average Ly- α effective optical depth after excluding damped absorption systems – are rare objects, only accounting for 0.1 per cent of all $15h^{-1}$ cMpc spectral segments drawn from the models. They probe a wide range in mass overdensity, δ_m , including underdense regions on $15h^{-1}$ cMpc scales, and so do not uniquely trace significant mass overdensities.
- Protoclusters with $M_{z=0} \simeq 10^{14} M_{\odot}$ exhibit a broad range of signatures in Ly- α absorption, with $\delta_{\tau_{\text{eff}}}$ ranging from -0.5 to > 8 . However, the vast majority (84 per cent) of sight lines passing through what we define as protoclusters in the simulations contain only low column density Ly- α forest absorption and have $\delta_{\tau_{\text{eff}}} < 1$. This signature is

identical to the field, so the majority of sight lines through protoclusters do not bear a telltale signature in Ly- α absorption. Most sight lines with high $\delta_{\tau_{\text{eff}}}$ are a result of passing through a damped absorption system. A small subset of sight lines through protoclusters do, however, exhibit high $\delta_{\tau_{\text{eff}}}$ due to coherently strong intergalactic Ly- α absorption systems i.e. CoSLAs. LLSs and high column density Ly- α forest absorbers are typically responsible for this absorption.

- Assuming the perfect removal of damped Ly- α absorption systems, CoSLAs are a good but non-unique probe of protoclusters at $z \simeq 2.4$. In the Sherwood simulation, approximately half of CoSLAs with $\delta_{\tau_{\text{eff}}} > 3.5$ trace protoclusters with $10^{14} \leq M_{z=0}/M_{\odot} \leq 10^{14.7}$. The other 46 per cent of CoSLAs arise from LLSs that are aligned along the line of sight.

- We find that threshold of $\delta_{\tau_{\text{eff}}} > 4.1$ – corresponding to regions on $15h^{-1}$ cMpc scales with 5.1 times the average Ly- α effective optical depth after excluding damped absorption systems – enables us to select a completely pure sample of protoclusters from simulated spectra. However, CoSLAs are rare within the volumes that protoclusters occupy: less than 0.1 per cent per cent of sight lines that pass through at least $5h^{-1}$ cMpc of a protocluster volume exhibit $\delta_{\tau_{\text{eff}}} > 4.1$, excluding absorption caused by damped systems. This means that any sample of protoclusters selected with the CoSLA technique will be incomplete. We stress, however, that throughout this work we are limited by the box size of the hydrodynamical simulations. In particular, there are no $10^{15} M_{\odot}$ cluster progenitors in any of the models considered here; it is possible that more massive protoclusters have a stronger association with CoSLAs.

Finally, we note that visual inspection of CoSLAs suggests that coherent Ly- α absorption typically selects structure orientated along the line of sight to the observer, regardless of whether or not this is associated with a protocluster. With the advent of large spectroscopic QSO surveys such DESI (Vargas-Magana et al. 2019) and WEAVE-QSO (Pieri et al. 2016) in the next few years, further investigation of the potential of CoSLAs for identifying intergalactic filaments in the high redshift Universe may be a worthwhile endeavour.

ACKNOWLEDGEMENTS

We thank Zheng Cai for comments on a draft version of this paper. The Sherwood simulations were performed with supercomputer time awarded by the Partnership for Advanced Computing in Europe (PRACE) 8th Call. We acknowledge PRACE for awarding us access to the Curie supercomputer, based in France at the Tre Grand Centre de Calcul (TGCC). This work also made use of the DiRAC Data Analytic system at the University of Cambridge, operated by the University of Cambridge High Performance Computing Service on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant (ST/K001590/1), STFC capital grants ST/H008861/1 and ST/H00887X/1, and STFC DiRAC Operations grant ST/K00333X/1. DiRAC is part of the National E-Infrastructure. We thank Volker Springel for making P-GADGET-3 available, and Simeon Bird for helpful advice regarding the incidence of DLAs in Illustris. We acknowledge the Virgo Consortium

for making their simulation data available. The EAGLE simulations were performed using the DiRAC-2 facility at Durham, managed by the ICC, and the PRACE facility Curie in France at TGCC. We also thank the Illustris collaboration for making their data publicly available. JSAM is supported by an STFC postgraduate studentship. JSB acknowledges the support of a Royal Society University Research Fellowship.

REFERENCES

- Adams S. M., Martini P., Croxall K. V., Overzier R. A., Silverman J. D., 2015, *Monthly Notices of the Royal Astronomical Society*, 448, 1335
- Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, *The Astrophysical Journal*, 584, 45
- Altay G., Theuns T., Schaye J., Crighton N. H., Dalla Vecchia C., 2011, *Astrophysical Journal Letters*, 737, 37
- Altay G., Theuns T., Schaye J., Booth C. M., Vecchia C. D., 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 2689
- Bădescu T., Yang Y., Bertoldi F., Zabludoff A., Karim A., Mag-nelli B., 2017, *The Astrophysical Journal*, 845, 172
- Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L., 2011, *Monthly Notices of the Royal Astronomical Society*, 410, 1096
- Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, *Monthly Notices of the Royal Astronomical Society*, 430, 2067
- Bird S., Vogelsberger M., Haehnelt M., Sijacki D., Genel S., Torrey P., Springel V., Hernquist L., 2014, *Monthly Notices of the Royal Astronomical Society*, 445, 2313
- Bolton J. S., Becker G. D., 2009, Resolving the high redshift Ly α forest in smoothed particle hydrodynamics simulations, doi:10.1111/j.1745-3933.2009.00700.x, <https://academic.oup.com/mnrasl/article-abstract/398/1/L26/1039975>
- Bolton J. S., Haehnelt M. G., Viel M., Springel V., 2005, *Monthly Notices of the Royal Astronomical Society*, 357, 1178
- Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T. S., Meiksin A., Regan J. A., Viel M., 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 897
- Booth C. M., Schaye J., 2009, *Monthly Notices of the Royal Astronomical Society*, 398, 53
- Cai Z., et al., 2016, *The Astrophysical Journal*, 833, 135
- Cai Z., et al., 2017a, *The Astrophysical Journal*, 837, 71
- Cai Z., et al., 2017b, *The Astrophysical Journal*, 839, 131
- Capak P. L., et al., 2011, *Nature*, 470, 233
- Chabrier G., 2003, *Publications of the Astronomical Society of the Pacific*, 115, 763
- Chiang Y. K., Overzier R., Gebhardt K., 2013, *Astrophysical Journal*, 779, 127
- Chiang Y. K., Overzier R., Gebhardt K., 2014, *The Astrophysical Journal Letters*, 782
- Chiang Y. K., et al., 2015, *Astrophysical Journal*, 808
- Cooke E. A., Hatch N. A., Muldrew S. I., Rigby E. E., Kurk J. D., 2014, *Monthly Notices of the Royal Astronomical Society*, 440, 3262
- Crain R. A., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 450, 1937
- Crain R. A., et al., 2017, *Monthly Notices of the Royal Astronomical Society*, 464, 4204
- Cucciati O., et al., 2014, *Astronomy & Astrophysics*, 570, A16
- Daddi E., et al., 2009, *Astrophysical Journal*, 694, 1517
- Dalla Vecchia C., Schaye J., 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 140
- Dawson K. S., et al., 2013, *Astronomical Journal*, 145, 10
- Faucher-Giguère C. A., Lidz A., Zaldarriaga M., Hernquist L., 2009, *Astrophysical Journal*, 703, 1416

- Finley H., Petitjean P., Noterdaeme P., Pâris I., 2014, *Astronomy & Astrophysics*, 572, 31
- Francis P. J., Hewett P. C., 1993, *The Astronomical Journal*, 105, 1633
- Francis P., et al., 1996, *The Astrophysical Journal*, 457, 490
- Franx M., Illingworth G., de Zeeuw T., 1991, *The Astrophysical Journal*, 383, 112
- Gladders M. D., Yee H. K. C., 2000, Technical report, A New Method For Galaxy Cluster Detection I: The Algorithm, <http://arxiv.org/abs/astro-ph/0004092><http://dx.doi.org/10.1086/301557>, doi:10.1086/301557, <http://arxiv.org/abs/astro-ph/0004092><http://dx.doi.org/10.1086/301557>
- Gurvich A., Burkhart B., Bird S., 2017, *The Astrophysical Journal*, 835, 175
- Haardt F., Madau P., 2001, [10.1038/nmat2978](https://arxiv.org/abs/10.1038/nmat2978)
- Haardt F., Madau P., 2012, *Astrophysical Journal*, 746
- Hatch N. A., et al., 2011, *Monthly Notices of the Royal Astronomical Society*, 410, 1537
- Hinshaw G., et al., 2013, *Astrophysical Journal, Supplement Series*, 208, 19
- Kim T.-S., Partl A. M., Carswell R. F., Müller V., 2013, *Astronomy & Astrophysics*, 552, 77
- Kravtsov A., Borgani S., 2012, *Annual Review of Astronomy and Astrophysics*, 50, 353
- Krolewski A., et al., 2018, *The Astrophysical Journal*, 861, 60
- Le Fèvre O., Deltorn J. M., Crampton D., Dickinson M., 1996, *The Astrophysical Journal*, 471, L11
- Lee K. G., Hennawi J. F., White M., Croft R. A., Ozbek M., 2014, *Astrophysical Journal*, 788, 49
- Lee K.-G., et al., 2016, *The Astrophysical Journal*, 817, 160
- Lee K.-G., et al., 2018, *The Astronomical Journal Supplement Series*, 238
- Lemaux B. C., et al., 2014, *Astronomy & Astrophysics*, 572, A41
- Lemaux B. C., et al., 2017, *Astrophysics A&A*, 615, 77
- Lovell C. C., Thomas P. A., Wilkins S. M., 2018, *Monthly Notices of the Royal Astronomical Society*, 474, 4612
- Lukić Z., Stark C. W., Nugent P., White M., Meiksin A. A., Almgren A., 2014, *Monthly Notices of the Royal Astronomical Society*, 446, 3697
- Madau P., Dickinson M., 2014, *Annual Review of Astronomy and Astrophysics*, 52, 415
- McAlpine S., et al., 2016, *Astronomy and Computing*, 15, 72
- Monaghan J. J., Lattanzio J. C., 1985, *Astronomy and Astrophysics*, 149, 135
- Morselli L., et al., 2014, *Astronomy & Astrophysics*, 568, A1
- Mukae S., et al., 2017, *The Astrophysical Journal*, 835, 281
- Muldrew S. I., Hatch N. A., Cooke E. A., 2015, *Monthly Notices of the Royal Astronomical Society*, 452, 2528
- Nelson D., et al., 2015, *Astronomy and Computing*, 13, 12
- Noterdaeme P., et al., 2012, *Astronomy & Astrophysics*, 547, L1
- Overzier R. A., 2016, *The Astronomy and Astrophysics Review*, 24, 14
- Peirani S., Weinberg D. H., Colombi S., Blaizot J., Dubois Y., Pichon C., 2014, *The Astrophysical Journal*, 784, 11
- Pieri M. M., et al., 2016, Technical report, WEAVE-QSO: A Massive Intergalactic Medium Survey for the William Herschel Telescope, <https://arxiv.org/pdf/1611.09388.pdf>. <https://arxiv.org/pdf/1611.09388.pdf>
- Planck Collaboration et al., 2014, *Astronomy & Astrophysics*, 571, A16
- Prochaska J. X., Wolfe A. M., 2009, *The Astrophysical Journal*, 696, 1543
- Puchwein E., Springel V., 2013, *Monthly Notices of the Royal Astronomical Society*, 428, 2966
- Rahmati A., Pawlik A. H., Raićević M., Schaye J., 2013, *Monthly Notices of the Royal Astronomical Society*, 430, 2427
- Schaye J., Dalla Vecchia C., 2008, *Monthly Notices of the Royal Astronomical Society*, 383, 1210
- Schaye J., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 446, 521
- Scoville N., et al., 2007, *The Astrophysical Journal Supplement Series*, 172, 1
- Springel V., 2005, The cosmological simulation code GADGET-2, doi:10.1111/j.1365-2966.2005.09655.x, <https://academic.oup.com/mnras/article-abstract/364/4/1105/1042826>
- Springel V., 2010, *Monthly Notices of the Royal Astronomical Society*, 401, 791
- Springel V., Hernquist L., 2003, *Monthly Notices of the Royal Astronomical Society*, 339, 289
- Springel V., et al., 2005, *Nature*, 435, 629
- Stark C. W., White M., Lee K. G., Hennawi J. F., 2015, *Monthly Notices of the Royal Astronomical Society*, 453, 311
- Steidel C. C., Adelberger K. L., Shapley A. E., Erb D. K., Reddy N. A., Pettini M., 2005, *The Astrophysical Journal*, 626, 44
- Tepper-García T., 2006, *Monthly Notices of the Royal Astronomical Society*, 369, 2025
- Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, *Monthly Notices of the Royal Astronomical Society*, 301, 478
- Toshikawa J., et al., 2018, *Publications of the Astronomical Society of Japan*, 70, 12
- Vargas-Magana M., Brooks D. D., Levi M. M., Tarle G. G., 2019
- Venemans B. P., et al., 2007, *A&A*, 461, 823
- Viel M., Haehnelt M. G., Springel V., 2004, *Monthly Notices of the Royal Astronomical Society*, 354, 684
- Villaescusa-Navarro F., et al., 2018, *The Astrophysical Journal*, 866, 135
- Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *Monthly Notices of the Royal Astronomical Society*, 436, 3031
- Vogelsberger M., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 444, 1518
- Wendland H., 1995, *Advances in Computational Mathematics*, 4, 389
- White M., 2000, *Astronomy & Astrophysics*, 367, 27
- Wold M., Armus L., Neugebauer G., Jarrett T. H., Lehnert M. D., 2003, *The Astronomical Journal*, 126, 1776
- Wylezalek D., et al., 2013, *Astrophysical Journal*, 769, 79
- Zheng W., et al., 2006, *The Astrophysical Journal*, 640, 574

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.